

Wed Mar 2 14:27:17 2005

```

Db          33 PHPOFG 39

RESULT 6
ID 092PK7; PRELIMINARY; PRT; 225 AA.
AC 092PK7;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein SWC00528.
GN ORFNames=SWC00528;
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiales; Sinorhizobium/Ensifer group; Sinorhizobium.
OX NCBI_TaxID=382;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396507; PubMed=11481430; DOI=10.1073/pnas.161294398;
RA Capela D., Barloy-Hubler F., Gouzy J., Bothe G., Ampe F., Batut J.,
RA Boleard F., Becker A., Boutry M., Cadieu E., Dreano S., Gloux S.,
RA Godrie T., Goffeau A., Kahn D., Kles A., Lelaure V., Masny D.,
RA Pohl T., Pottecielie D., Puehler A., Purnelle B., Ramsperger U.,
RA Renard C., Thebaud P., Vandenbol M., Weidner S., Galibert F.;
RT "Analysis of the chromosome sequence of the legume symbiont
RT Sinorhizobium meliloti strain 1021."
RL Proc. Natl. Acad. Sci. U.S.A. 98:9877-9882(2001).
DR EMBL; AL591788; CAC46316.1; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR InterPro; IPR000379; Ser_estrs.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 225 AA; 25011 MW; C87CD753E13D92B CRC64;

Query Match          88.6%; Score 39; DB 2; Length 225;
Best Local Similarity 85.7%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy          2 XHPORFG 8
Db          33 PHPOFG 39

RESULT 7
ID 08UED4; PRELIMINARY; PRT; 225 AA.
AC 08UED4; Q7CYF9;
DT 01-JUN-2002 (TREMBLrel. 21, Created)
DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)
DT 25-OCT-2004 (TREMBLrel. 28, Last annotation update)
DE Hypothetical protein Atu1826 (AGR_C_3351P).
GN OrderedLocustNames=AGR_C_3351, Atu1826;
OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Rhizobiales; Rhizobium/Agrobacterium group; Agrobacterium.
OX NCBI_TaxID=176299;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Dupont;
RX MEDLINE=21608551; PubMed=11743193; DOI=10.1126/science.1066804;
RA Wood D.W., Seubel J.C., Kaul R., Monks D.E., Kitaajima J.P.,
RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Bovee D. St.,
RA Chapman P., Clendinning J., Deatherage G., Gillet W., Grant C.,
RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
RA Raymond C., Rouse G., Saenphimmach C., Wu Z., Romero P., Gordon D.,
RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan M., Perry M.,
RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
RA Neester E.W.;
RT "The genome of the natural genetic engineer Agrobacterium tumefaciens
RT C58."
RL Science 294:2317-2323(2001).

[2]
RN SEQUENCE FROM N.A.
RC STRAIN=Cereon;
RX MEDLINE=21608551; PubMed=11743194; DOI=10.1126/science.1066803;
RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
RA Qurollo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
RA Houtel K., Gordon J., Vaudin M., Iartchouk O., Epp A., Liu F.,
RA William C., Allinger M., Doughny D., Scott C., Lappas C., Markelz B.,
RA Flanagan C., Crowell C., Gironson V., Lomo C., Sear C., Strub G.,
RA Cleio C., Slater S.;
RT "Genome sequence of the plant pathogen and biotechnology agent
RT Agrobacterium tumefaciens C58."
RL Science 294:2323-2328(2001).
DR EMBL; AE009137; AA142822.1; -.
DR EMBL; AE008102; AAK87595.1; -.
DR PIR; AH2800; AH2800.
DR PIR; B97580; B97580.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR InterPro; IPR000379; Ser_estrs.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 225 AA; 25084 MW; 1B7BDF5600841833 CRC64;

Query Match          88.6%; Score 39; DB 2; Length 225;
Best Local Similarity 85.7%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy          2 XHPORFG 8
Db          33 PHPOFG 39

RESULT 8
ID 06FZW3; PRELIMINARY; PRT; 226 AA.
AC 06FZW3;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocustNames=BQ059310;
OS Bartonella quintana (Rochalimaea quintana).
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bartonellaceae; Bartonella.
OX NCBI_TaxID=803;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=Toulouse;
RX PubMed=15210978; DOI=10.1073/pnas.0305659101;
RA Alenmark U.C.M., Frank A.C., Karlberg E.O., Legault B.-A., Ardell D.H.,
RA Cabaeck B., Eriksson A.-S., Naestlund A.K., Handley S.A., Huvel M.,
RA La Scolia B., Holmberg M., Andersson S.G.B.;
RT "The louse-borne human pathogen Bartonella quintana is a genomic
RT derivative of the zoonotic agent Bartonella henselae."
RL Proc. Natl. Acad. Sci. U.S.A. 101:9716-9721(2004).
DR EMBL; BX897700; CAP26085.1; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR InterPro; IPR000379; Ser_estrs.
KW Complete proteome.
SQ SEQUENCE 226 AA; 25373 MW; D2AD7C07BEC2B625 CRC64;

Query Match          88.6%; Score 39; DB 2; Length 226;
Best Local Similarity 85.7%; Pred. No. 24;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy          2 XHPORFG 8
Db          33 PHPOFG 39

RESULT 9
ID 06G3B7; PRELIMINARY; PRT; 226 AA.
AC 06G3B7;
SQ SEQUENCE 294:2317-2323(2001).

```

DT 05-JUN-2004 (TReMBLrel. 27, Created)  
 DT 05-JUN-2004 (TReMBLrel. 27, Last sequence update)  
 DT 05-JUN-2004 (TReMBLrel. 27, Last annotation update)  
 DE Hypothetical protein.  
 GN OrderedLocusNames=BH08660;  
 OS Bartonella henselae (Rochalimaea henselae).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 CC Bartonellaceae; Bartonella.  
 NCBI\_TaxID=38323;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 49882 / Houston 1;  
 RX PubMed=15210978; DOI=10.1073/pnas.0305659101;  
 RA Almark U.C.M., Frank A.C., Karlberg E.O., Legault B.-A., Ardeli D.H.,  
 RA Candaeck B., Eriksson A.-S., Naeslund A.K., Handley S.A., Huet M.,  
 RA La Scoia B., Holmberg M., Andersson S.G.E.;  
 RT "The house-borne human pathogen Bartonella quintana is a genomic  
 RT derivative of the zoonotic agent Bartonella henselae.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 101:9716-9721(2004).  
 DR EMBL: BX897699; CAF27664.1; -;  
 DR GO:GO:0003824; F:catalytic activity; IEA.  
 DR InterPro: IPR000379; Set\_sestr.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 226 AA; 25230 MW; 7216738444690A60 CRC64;

Query Match 88.6%; Score 39; DB 2; Length 226;  
 Best Local Similarity 85.7%; Pred. No. 24;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8  
 :|||||  
 DB 33 PHPQFG 39

RESULT 10  
 Q98NR9 PRELIMINARY; PRT; 228 AA.

DT 01-OCT-2001 (TReMBLrel. 18, Created)  
 DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)  
 DT 01-OCT-2001 (TReMBLrel. 25, Last annotation update)  
 DE M110014 protein.  
 GN OrderedLocusNames=m110014;  
 OS Rhizobium loti (Mesorhizobium loti).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 CC Rhizobiaceae; Mesorhizobium.  
 NCBI\_TaxID=381;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asanizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Ideawata K., Iehikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,  
 RA Takuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti.";  
 RL DNA Res. 7:331-338(2000).  
 DR EMBL: AP002994; BAB47692.1; -;  
 DR GO:GO:0003824; F:catalytic activity; IEA.  
 DR InterPro: IPR000379; Set\_sestr.  
 KW Complete proteome.  
 SQ SEQUENCE 228 AA; 25544 MW; B1A116C8EAE893BC CRC64;

Query Match 88.6%; Score 39; DB 2; Length 228;  
 Best Local Similarity 85.7%; Pred. No. 24;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8  
 :|||||  
 DB 37 PHPQFG 43

RESULT 11  
 Q96UF7 PRELIMINARY; PRT; 330 AA.

DT 01-DEC-2001 (TReMBLrel. 19, Created)  
 DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)  
 DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
 DE Subtilisin-like serine protease PR1B (Fragment).  
 GN Name=PR1B;  
 OS Metarhizium anisopliae var. acridum.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreomycetidae; Ascomycetes; Clavicipitaceae;  
 CC Microsporid Clavicipitaceae; Metarhizium.  
 NCBI\_TaxID=92637;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FI-985;  
 RA Bagga S., St. Leger R.T.;  
 RL Submitted (OCT-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AJ416582; CAC95046.1; -;  
 DR HSP; P06873; 2PRK.  
 DR GO:GO:0008233; F:peptidase activity; IEA.  
 DR GO:GO:0004289; F:subtilisin activity; IEA.  
 DR GO:GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro: IPR000209; Pept\_S8\_S53.  
 DR InterPro: IPR009020; Prot\_inh\_S8A.  
 DR InterPro: IPR010259; Prot\_inh\_S8A.  
 DR Pfam: PF00082; Peptidase\_S8; 1.  
 DR Pfam: PF05922; Subtilisin\_N; 1.  
 DR PRINTS: PR00723; SUBTILISIN.  
 DR PROSITE: PS00136; SUBTILASE\_ASP; 1.  
 KW Protease.  
 FT NON\_TER 330 330  
 SQ SEQUENCE 330 AA; 34673 MW; 09E294C2864E9130 CRC64;

Query Match 88.6%; Score 39; DB 2; Length 330;  
 Best Local Similarity 85.7%; Pred. No. 36;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8  
 :|||||  
 DB 149 SHPQFG 155

RESULT 12  
 P511\_SCHPO STANDARD; PRT; 379 AA.

DT 01-FEB-1996 (Rel. 33, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 05-JUN-2004 (Rel. 44, Last annotation update)  
 DE Protein p511 (Protein p51).  
 GN Name=p511; Synonyms=p51; ORFNames=SPCC830.07c;  
 OS Schizosaccharomyces pombe (Fission yeast).  
 OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
 CC Schizosaccharomycetes; Schizosaccharomycetaceae;  
 CC Schizosaccharomycetes.  
 NCBI\_TaxID=4896;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=95290501; PubMed=7772606; DOI=10.1016/0167-4781(95)00063-M;  
 RA Park S.-K., Chon S.-K., Yoo H.-S.;  
 RT "A cDNA of Schizosaccharomyces pombe encoding a homologue of DnaJ-like  
 RT protein.";  
 RL Biochim. Biophys. Acta 1262:87-90(1995).  
 RN (2)  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=972;  
 RX MEDLINE=21648401; PubMed=11859360; DOI=10.1038/nature724;  
 RA Wood V., Gwilliam R., Rajandream M.A., Lyne M., Lyne R., Stewart A.,  
 RA Sgouras J., Peat N., Hayles J., Baker S., Basham D., Bowman S.,  
 RA Brooks K., Brown D., Brown S., Chillingworth T., Churcher C.M.,

RA Collins M., Connor R., Cronin A., Davis P., Fellwell T., Fraser A.,  
 RA Gentles S., Goble A., Hamlin N., Harris D., Hidalgo J., Hodgson G.,  
 RA Holroyd S., Hornsby T., Howarth S., Huckle E.J., Hunt S., Jagels K.,  
 RA James K., Jones L., Jones M., Leather S., McDonald S., McLean J.,  
 RA Mooney P., Moule S., Mungall K., Murphy L., Niblett D., Odell C.,  
 RA Oliver K., O'Neill S., Pearson D., Quail M.A., Rabbittowlesch E.,  
 RA Rutherford K., Rutter S., Saunders D., Seeger K., Sharp S.,  
 RA Skelton J., Simmonds M., Squares R., Squares S., Stevens K.,  
 RA Taylor K., Taylor R.G., Tivey A., Walsh S.V., Warren T., Whitehead S.,  
 RA Woodward J., Volckaert G., Aert R., Robben J., Grynoprez B.,  
 RA Wellens I., Vanstreels E., Rieger M., Schaefer M., Mueller-Auer S.,  
 RA Gabel C., Fuchs M., Fritze C., Holzer E., Moser D., Hilbert H.,  
 RA Borzym K., Langer I., Beck A., Lehnrich H., Reinhardt R., Pohl T.M.,  
 RA Eger P., Zimmermann W., Medler H., Wandt R., Purnelle B.,  
 RA Goffeau A., Cadieu E., Dreano S., Gloux S., Delaure V., Notter S.,  
 RA Galbert F., Aves S.J., Xiang Z., Hunt C., Moore K., Hurst S.M.,  
 RA Lucas M., Rochet M., Galliard C., Tallada V.A., Garçon A., Thode G.,  
 RA Daga R.R., Cruzado L., Jimenez J., Sanchez M., del Rey F., Benito J.,  
 RA Dominguez A., Revuelta J.L., Moreno S., Armstrong J., Forburg S.L.,  
 RA Cerutti L., Lowe T., McCombie W.R., Paulsen I., Potashkin J.,  
 RA Shpakovskii G.V., Ussery D., Barrell B.G., Nurse P.,  
 RA "The genome sequence of Schizosaccharomyces pombe";  
 RL Nature 415:871-880(2002).  
 CC -!- FUNCTION: Required for nuclear migration during mitosis. It is  
 CC required for the normal initiation of translation.  
 CC -!- SIMILARITY: Contains 1 J domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL, L37753; AAA74732.1; -.  
 DR EMBL, AL109850; CAB52880.1; -.  
 DR PIR, S55900; S55900.  
 DR PIR, T11633; T11633.  
 DR HSSP, P25685; 1HDJ.  
 DR GeneDB, SPombe; SPCC830.07c; -.  
 DR InterPro; IPR002939; DnaJ\_C.  
 DR InterPro; IPR001623; DnaJ\_N.  
 DR InterPro; IPR008971; HSP40\_DnaJ\_pcp.  
 DR InterPro; IPR003095; Hsp\_DnaJ.  
 DR Pfam; PF00226; DnaJ\_1.  
 DR Pfam; PF01556; DnaJ\_C\_1.  
 DR PRINTS; PR00625; DNAJPROTEIN.  
 DR SMART; SM00371; DnaJ\_1.  
 DR PROSITE; PS00636; DnaJ\_1; 1.  
 DR PROSITE; PS0076; DnaJ\_2; 1.  
 DR Cell cycle; Chaperone; DNA-binding.  
 FT DOMAIN 1 70 J-domain.  
 FT DOMAIN 80 182 Gly-rich.  
 FT CONFLICT 91 91 G -> C (in Ref. 1).  
 SQ SEQUENCE 379 AA; 40260 MW; B8B349A1903F988A CRC64;

Query Match 88.6%; Score 39; DB 1; Length 379;  
 Best Local Similarity 75.0%; Pred. No. 41;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RXHPQFG 8  
 Db 187 RSHPSFG 194

RESULT 13  
 ID 096UP9 PRELIMINARY; PRT; 385 AA.  
 AC 096UP9;  
 DT 01-DEC-2001 (TREMBlrel. 19, Created)  
 DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)

DE Subtilisin-like protease PR1B.  
 GN Name=PR1B;  
 OS Metarhizium anisopliae var. anisopliae.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreomycetidae; Hypocreales; Clavicipitaceae;  
 OC mitosporic Clavicipitaceae; Metarhizium.  
 OX NCB1\_TaxID=92636;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ARSEF 820;  
 RA Bagga S., St Leger R.J.;  
 RL Submitted (OCT-2001) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AJ41690; CAC95044.1; -.  
 DR HSSP; P06873; 1IC6.  
 DR GO; GO:000823; F:peptidase activity; IEA.  
 DR GO; GO:0004289; F:subtilase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000183; Decarboxylase.  
 DR InterPro; IPR009020; Pept\_S8\_553.  
 DR InterPro; IPR009020; Pept\_inh\_propept.  
 DR InterPro; IPR010259; Pept\_inh\_S8A.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR Pfam; PF05922; Subtilisin\_N; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR PROSITE; PS00879; ODR\_DC\_2\_2; UNKNOWN\_1.  
 DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
 KW Protease.  
 SQ SEQUENCE 385 AA; 40094 MW; DB31BA0DC3A29FC3 CRC64;

Query Match 88.6%; Score 39; DB 2; Length 385;  
 Best Local Similarity 85.7%; Pred. No. 42;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPQFG 8  
 Db 149 THPQFG 155

RESULT 14  
 ID 09P3Y1 PRELIMINARY; PRT; 385 AA.  
 AC 09P3Y1;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Subtilisin-like protease PR1B.  
 GN Name=PR1B;  
 OS Metarhizium anisopliae var. anisopliae.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreomycetidae; Hypocreales; Clavicipitaceae;  
 OC mitosporic Clavicipitaceae; Metarhizium.  
 OX NCB1\_TaxID=92636;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ARSEF 2575;  
 RA Bagga S., Screen S.E., St Leger R.J.;  
 RL Submitted (JUN-2000) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; AJ289823; CAB95012.1; -.  
 DR HSSP; P06873; 1IC6.  
 DR GO; GO:000823; F:peptidase activity; IEA.  
 DR GO; GO:0004289; F:subtilase activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000183; Decarboxylase.  
 DR InterPro; IPR009020; Pept\_S8\_553.  
 DR InterPro; IPR009020; Pept\_inh\_propept.  
 DR InterPro; IPR010259; Pept\_inh\_S8A.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR Pfam; PF05922; Subtilisin\_N; 1.  
 DR PRINTS; PR00723; SUBTILISIN.  
 DR PROSITE; PS00879; ODR\_DC\_2\_2; UNKNOWN\_1.  
 DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
 KW Protease.  
 SQ SEQUENCE 385 AA; 39959 MW; D9857DBEFC25D7C CRC64;

Query Match 88.6%; Score 39; DB 2; Length 385;  
 Best Local Similarity 85.7%; Pred. No. 42;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8  
 :|||||  
 Db 149 THPQFG 155

## RESULT 15

01410 PRELIMINARY; PRT; 386 AA.  
 ID 014410  
 AC 014410;  
 DT 01-JAN-1998 (TrEMBLrel. 05, Created)  
 DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Subtilisin-like protease PriB (Fragment).  
 GN Name=PriB;  
 OS Metarhizium anisopliae.  
 OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 OC Hypocreomycetidae; Hypocreales; Clavicipitaceae;  
 OC Mitosporic Clavicipitaceae; Metarhizium.  
 OX NCBI\_TaxID=5530;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ARSEF 2575;  
 RK MEDLINE=97473490; PubMed=9332344; DOI=10.1016/S0378-1119(97)00132-7;  
 RA Joshi L., St Leger R.J., Roberts D.W.;  
 RT "Isolation of a cDNA encoding a novel subtilisin-like protease (PriB)  
 RT from the entomopathogenic fungus, Metarhizium anisopliae using  
 RT differential display-RT-PCR.";  
 RL Gene 197:1-8(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=ARSEF 2575;  
 RA Joshi L.S.T., Leger R.J., Roberts D.W.;  
 RL Submitted (MAY-1996) to the EMBL/Genbank/DBJ databases.  
 DR EMBL; U59484; AAC9831.1; -.  
 DR HSP; P06873; 2PRK.  
 DR GO; GO:0008233; F:peptidase activity; IEA.  
 DR GO; GO:0004289; F:subtilisin-like protease activity; IEA.  
 DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
 DR InterPro; IPR000183; Decarboxylase2.  
 DR InterPro; IPR000209; Pept\_S8\_S53.  
 DR InterPro; IPR009020; Prot\_inh\_propept.  
 DR InterPro; IPR010259; Prot\_inh\_S8A.  
 DR Pfam; PF00082; Peptidase\_S8; 1.  
 DR Pfam; PF05922; Subtilisin\_N; 1.  
 DR PRINTS; PRO0723; SUBTILISIN.  
 DR PROSITE; PS00879; ODR\_DC\_2\_2; UNKNOWN\_1.  
 DR PROSITE; PS00136; SUBTILASE\_ASP; 1.  
 KW Protease.  
 FT NON\_TER 386  
 FT 386  
 SQ SEQUENCE 386 AA; 40031 MW; E805E0751C471568 CRC64;

Query Match 88.6%; Score 39; DB 2; Length 386;  
 Best Local Similarity 85.7%; Pred. No. 42;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8  
 :|||||  
 Db 150 THPQFG 156

Search completed: March 2, 2005, 12:44:18  
 Job time : 30.7236 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 36.0325 Seconds  
(without alignments)  
85.869 Million cell updates/sec

Title: SEQ7  
Perfect score: 44  
Sequence: 1 rxhpqfsg 8

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq16Dec04:\*

1: geneseqp1980s:\*\n2: geneseqp1990s:\*\n3: geneseqp2000s:\*\n4: geneseqp2001s:\*\n5: geneseqp2002s:\*\n6: geneseqp2003as:\*\n7: geneseqp2003bs:\*\n8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	88.6	8	6	ABP60360
2	39	88.6	9	2	AAW32485
3	39	88.6	9	2	AAW59211
4	39	88.6	9	2	AAW44010
5	39	88.6	9	2	AAW06914
6	39	88.6	9	2	AAW0795
7	39	88.6	9	5	ABG31053
8	39	88.6	9	5	AAW78476
9	39	88.6	9	6	AAO16086
10	39	88.6	9	6	ABP55547
11	39	88.6	9	6	AAE37229
12	39	88.6	9	6	AAE32860
13	39	88.6	9	6	AAE33270
14	39	88.6	9	6	ABG72479
15	39	88.6	9	7	ADB84587
16	39	88.6	9	7	ADJ34695
17	39	88.6	9	8	ADJ36219
18	39	88.6	9	8	ADJ88138
19	39	88.6	9	8	ADN29604
20	39	88.6	9	8	ADN82347
21	39	88.6	10	2	AAW64446
22	39	88.6	10	2	AAW50025
23	39	88.6	10	3	AAW57325
24	39	88.6	10	3	AAW67424
25	39	88.6	10	3	AAW19071

26	39	88.6	10	4	AAW61168	AAW61168	Paramagne
27	39	88.6	10	4	AAW99026	AAW99026	Streptavi
28	39	88.6	10	4	AAW02084	AAW02084	Streptavi
29	39	88.6	10	5	AAO14759	AAO14759	Human Int
30	39	88.6	10	7	ADD29929	ADD29929	Antibody-
31	39	88.6	11	2	AAW52691	AAW52691	PASK46-p1
32	39	88.6	11	2	AAW52692	AAW52692	PASK46-p1
33	39	88.6	12	2	AAW52696	AAW52696	PASK46-p1
34	39	88.6	12	2	AAW81821	AAW81821	Streptavi
35	39	88.6	18	5	AAW78473	AAW78473	Linker co
36	39	88.6	20	6	AAW37002	AAW37002	Artificia
37	39	88.6	23	5	AAO19984	AAO19984	Peptide o
38	39	88.6	42	5	AAO19988	AAO19988	Protein o
39	39	88.6	42	5	AAO19986	AAO19986	Protein o
40	39	88.6	46	8	ADP32686	ADP32686	Receptor
41	39	88.6	46	8	ADJ55658	ADJ55658	ED Peptid
42	39	88.6	47	2	AAW51892	AAW51892	VH1 Leade
43	39	88.6	47	5	ABW81776	ABW81776	Plasmid P
44	39	88.6	132	8	ADJ36263	ADJ36263	Self-coal
45	39	88.6	155	2	AAW32483	AAW32483	Kappa 11g

## ALIGNMENTS

RESULT 1	ABP60360	standard, peptide, 8 AA.
ID	ABP60360	standard, peptide, 8 AA.
AC	ABP60360	
XX		
XX		
DT	28-MAR-2003	(first entry)
XX		
DE	Streptavidin tag peptide SEQ ID NO 1.	
XX		
KW	Streptavidin; protein chip; microtitre plate; detection.	
XX		
OS	Synthetic.	
XX		
PN	DE10113776-A1.	
XX		
PD	02-OCT-2002.	
XX		
PF	21-MAR-2001; 2001DE-01013776.	
XX		
PR	21-MAR-2001; 2001DE-01013776.	
XX		
PA	(BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.	
XX		
PI	Schmidt T;	
XX		
DR	WPI; 2003-031166/03.	
XX		
PT	New isolated peptide, useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.	
XX		
PS	Disclosure; Page 2; 18pp; German.	
XX		
CC	The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides stronger binding than a single tag, but are displaced by a competitor. (I) does not interfere with the function of attached proteins (II) (so it may not be essential to remove it); facilitates detection and has easily controllable binding properties. (I) is particularly used for purifying FP from dilute solution in batch formats (which use simpler apparatus than column methods and result in lower loss of FP). The present sequence is that of a streptavidin tag peptide disclosed with the invention	

XX Sequence 8 AA;

Query Match 88.6%; Score 39; DB 6; Length 8;  
Best Local Similarity 85.7%; Pred. NO. 1.8e+06;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8  
DB 2 XHPQFG 8

RESULT 2  
AAW32485  
ID AAW32485 standard; peptide; 9 AA.

AC AAW32485;

DT 27-MAR-1998 (first entry)

DE Strep-tag peptide.

XX Catalytic antibody; growth factor; B-cell mitogenesis; LHL; L protein;  
XX hen egg lysozyme; strep-tag; purification.

XX Synthetic.

XX WO9735887-A1.

XX 02-OCT-1997.

XX 26-MAR-1997; 97WO-AU000194.

XX 26-MAR-1996; 96AU-00008951.

XX 27-FEB-1997; 97AU-00005375.

XX (AMRA-) AMRAD OPERATIONS PTY LTD.

XX Koentgen F., Suesse GM, Tarlinton DM, Treutlein HR;

XX WPI; 1997-489572/45.

XX New catalytic antibody precursors - comprising a B-cell surface molecule  
XX binding portion which can induce B-cell mitogenesis.

XX Example 3; Page 40; 109pp; English.

XX This peptide comprises the strep-tag peptide that is recognised by  
CC streptavidin. A form of novel growth factor LHL (see AAW32479) was  
CC generated by PCR that contains a FLAG epitope (see AAW32484) at its N-  
CC terminus and the strep-tag at its C-terminus. The construct is designated  
CC LHL.seq (see AAW32481), where L is the immunoglobulin binding entity from  
CC Peptostreptococcus magnus and H is residues 42-62 of hen egg lysozyme.  
CC The strep-tag was used for purification of LHL.seq over a streptavidin  
CC column. Thus, the LHL.seq was not purified on the basis of binding  
CC immunoglobulin, thereby eliminating potential contamination by other  
CC bacterial proteins which also bind immunoglobulins. LHL.seq has identical  
CC activity to that of LHL, and can be used in novel methods for the  
CC generation of catalytic antibodies

XX Sequence 9 AA;

Query Match 88.6%; Score 39; DB 2; Length 9;  
Best Local Similarity 85.7%; Pred. NO. 1.8e+06;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8  
DB 3 RHPQFG 9

RESULT 3  
AAW59211

ID AAW59211 standard; peptide; 9 AA.

AC AAW59211;

DT 27-AUG-1998 (first entry)

DE Streptavidin tagged peptide ligand #1.

XX Streptavidin; ligand; binding affinity; mutant; isolation; purification;  
XX recover; immobilise.

XX Synthetic.

XX EP835934-A2.

XX 15-APR-1998.

XX 09-OCT-1997; 97EP-00117504.

XX 10-OCT-1996; 96DE-01041876.

XX (BIOA-) INST BIOANALYTIK GMBH.

XX Skerra A, Voss S;

XX WPI; 1998-218868/20.

XX Streptavidin mutants with higher binding affinity for peptide ligands -  
XX have mutation in amino acid region 44-53, used to isolate, purify or  
XX determine fusion proteins including these ligands.

XX Claim 10; Page 11; 21pp; German.

XX AAW59211 and AAW59212 are ligands used in a method to assay binding  
CC affinity of streptavidin mutants. These mutants have a mutation within  
CC the amino acid (aa) region 44-53 of the wild-type protein show a higher  
CC binding affinity than the wild-type for peptide ligands that include the  
CC sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z  
CC are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant streptavidin  
CC mutants can be used to isolate, purify and determine proteins or to  
CC determine/recover substances that contain streptavidin-binding groups.  
CC Such compounds may also be used to immobilise fusions on microtitre  
CC plates, microbeads or sensor chips

XX Sequence 9 AA;

Query Match 88.6%; Score 39; DB 2; Length 9;  
Best Local Similarity 85.7%; Pred. NO. 1.8e+06;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8  
DB 3 RHPQFG 9

RESULT 4  
AAW44010  
ID AAW44010 standard; peptide; 9 AA.

AC AAW44010;

DT 05-JUN-1998 (first entry)

DE Strep peptide epitope used in an epitope tagged prion protein construct.

XX prion; epitope; FLAG; strep; poly-histidine; haemagglutinin; recombinant;  
XX transgenic animal; scrapie; Creutzfeldt-Jakob disease; CJD;  
XX bovine spongiform encephalopathy; BSE.

XX Synthetic.

XX WO9746572-A1.

```

PD 11-DEC-1997;
XX
XX 29-MAY-1997; 97WO-US009289.
XX
XX 06-JUN-1996; 96US-00660626.
XX
XX (REGC ) UNIV CALIFORNIA.
XX
XX Prusiner SB, Telling GC, Cohen FE, Scott MR;
XX
XX WPI; 1998-042112/04.
XX
XX Nucleic acid construct encoding biologically active protein and epitope -
XX especially epitope-tagged prion protein.
XX
XX Claim 4; Page 48; 62pp; English.
XX
XX This sequence represents an artificial Strep peptide epitope. It is used
XX in a recombinant nucleic acid construct encoding an epitope-tagged prion
XX protein (PrP). The construct comprises a first nucleic acid sequence
XX encoding an amino acid sequence of a biologically active protein fragment
XX and a second nucleic acid sequence encoding a heterologous epitope
XX domain. The heterologous epitope domain is a peptide selected from a
XX peptide group of FLAG, Strep, poly-histidine, human c-myc peptide
XX recognised by monoclonal antibody 9B10 and haemagglutinin peptide
XX recognised by monoclonal antibody 12CA5. The protein is a natural,
XX synthetic or chimeric PrP molecule. The protein has two different, three-
XX dimensional conformations and the epitope domain is spatially positioned
XX relative to the protein such that the epitope domain is more exposed in a
XX first conformation relative to a second conformation. The nucleic acid
XX construct may be used for the production of transgenic animals or cells
XX that are useful in a method for distinguishing between different
XX conformational shapes of a protein. These methods are particularly useful
XX in studying diseases caused by prion proteins, e.g. Creutzfeldt-Jakob
XX disease (CJD), scrapie and bovine spongiform encephalopathy (BSE)
XX
XX Sequence 9 AA;
XX
XX Query Match 88.6%; Score 39; DB 2; Length 9;
XX Best Local Similarity 85.7%; Pred. No. 1.8e+06;
XX Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2 XHPORFG 8
XX :|||||
XX Db 3 RHPORFG 9
XX
XX RESULT 5
XX ID AAY06914 strand; peptide; 9 AA.
XX
XX AAY06914;
XX
XX 01-JUL-1999 (first entry)
XX
XX Strep-tag sequence.
XX
XX Growth factor precursor; B-cell surface; T cell surface; CAB; hepatitis;
XX catalytic antibody; immunoglobulin; tumour necrosis factor; influenza;
XX rheumatoid arthritis; toxic shock syndrome; viral docking receptor;
XX human immune virus; tumour-specific antigen; amyloid plaque; myeloma;
XX Alzheimer's disease; IGE; allergy; asthma; drug detoxification;
XX autoimmune; inflammatory disease; gene therapy; strep-tag.
XX
XX Synthetic.
XX
XX WO915563-A1.
XX
XX 01-APR-1999.
XX
XX 18-SEP-1998; 98WO-AU000783.
XX
XX 19-SEP-1997; 97AU-00009306.

```

```

XX
XX (AMRA-) AMRAD OPERATIONS PTY LTD.
XX
XX Koentgen F, Sness GM, Tarlinton DM, Treutlein HR;
XX
XX WPI; 1999-244394/20.
XX
XX Growth factor precursor cleaved by antigen-specific catalytic antibody.
XX
XX Example 3; Page 42; 101pp; English.
XX
XX The invention relates to a growth factor precursor that comprises B-cell
XX surface binding part, T cell surface binding part, antigen cleavable by a
XX catalytic antibody (CAB); and a peptide comprising heavy and light chains
XX of immunoglobulin. When the antigen is cleaved the B cell surface part
XX can interact with its target. The growth factor precursors are used to
XX select B cells that produce Ag-specific CAB, and to generate CAB from
XX such cells (by inducing mitogenesis, caused by the growth factor released
XX by specific cleavage). The Ag-specific CAB can be directed against, e.g.
XX tumour necrosis factor (treatment of rheumatoid arthritis or toxic shock
XX syndrome); viral docking receptors (treatment of human immune virus,
XX hepatitis and influenza infections); tumour-specific antigens; amyloid
XX plaque (treatment of Alzheimer's disease or myeloma) or IGE (treatment of
XX allergies such as asthma). CAB may also be used for drug detoxification,
XX to treat autoimmune or inflammatory diseases and to eliminate
XX environmental or industrial pollutants, such as plastics and petroleum.
XX Particularly the growth factor precursors are produced by delivering the
XX corresponding nucleic acid in a viral or other gene therapy vector. The
XX present sequence represents a strep-tag
XX
XX Sequence 9 AA;
XX
XX Query Match 88.6%; Score 39; DB 2; Length 9;
XX Best Local Similarity 85.7%; Pred. No. 1.8e+06;
XX Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX QY 2 XHPORFG 8
XX :|||||
XX Db 3 RHPORFG 9
XX
XX RESULT 6
XX ID AAB30795 strand; peptide; 9 AA.
XX
XX AAB30795;
XX
XX 02-APR-2001 (first entry)
XX
XX Amino acid sequence of a Strep epitope.
XX
XX SCHAG; self-coalesce; higher-order aggregate; amyloidogenic domain;
XX aggregation; fibril; phenotypic alteration; gene therapy;
XX disease resistance; plant pigmentation; prion disease.
XX
XX Synthetic.
XX
XX WO200075324-A2.
XX
XX 14-DEC-2000.
XX
XX 09-JUN-2000; 2000WO-US015876.
XX
XX 09-JUN-1999; 99US-0138833P.
XX
XX (ARCH-) ARCH DEV CORP.
XX
XX Lindquist S, Li L, Ma J, Liu J, Sondheimer N, Scheibel T;
XX
XX WPI; 2001-061723/07.
XX
XX New nucleic acid encoding chimeric proteins with self-assembly
XX properties, useful e.g. for diagnosis and treatment of prion diseases,

```

PT also related aggregates, fibrils and polymers.  
 XX  
 PS Disclosure; Page 16; 188pp; English.  
 CC  
 CC The present sequence represents an epitope, which may be attached to  
 CC chimeric polypeptides of the invention. The specification describes  
 CC chimeric polypeptides which comprise at least one SCHAG (self-coalesces  
 CC into higher-order aggregates) amino acid sequence fused in frame with a  
 CC glutathione-S-transferase or a staphylococcal nuclear protein). The  
 CC specification also describes chimeric polypeptides that comprises an  
 CC amyloidogenic domain that causes aggregation into fibrils. The chimeric  
 CC polypeptides are used to prepare polymers with multiple reactivities,  
 CC e.g. derivatised with enzymes, or specific binding partners, and useful  
 CC for performing multi-step chemical reactions. They can be used  
 CC to create an inducible, or stable phenotypic alterations in a cell, e.g. for  
 CC gene therapy, protein production, imparting disease resistance to plants,  
 CC altering plant pigmentation and for diagnosis and treatment of prion  
 CC diseases  
 CC  
 SO Sequence 9 AA;  
 Query Match 88.6%; Score 39; DB 4; Length 9;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 XHPQFG 8  
 DB 3 RHPQFG 9  
 RESULT 7  
 ABG31053  
 ID ABG31053 standard; peptide; 9 AA.  
 XX  
 AC ABG31053;  
 XX  
 DT 05-NOV-2002 (first entry)  
 XX  
 DE Peptide production method associated strep-tag #1.  
 XX  
 KM Peptide production method; in vitro translation; in vitro transcription;  
 KM strep-tag.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200253582-A2.  
 XX  
 PD 11-JUL-2002.  
 XX  
 PF 06-DEC-2001; 2001WO-0P010682.  
 XX  
 PR 28-DEC-2000; 2000JP-00401417.  
 XX  
 PR 15-JAN-2001; 2001JP-00006910.  
 XX  
 PR 27-JUL-2001; 2001JP-00227094.  
 XX  
 PR 26-SEP-2001; 2001JP-00294795.  
 XX  
 PA (POST-) POST GENOME INST CO LTD.  
 XX  
 PI Inoue A, Shimizu Y, Ueda T;  
 XX  
 DR WPI; 2002-599608/64.  
 XX  
 PT Producing peptides, using reaction system for transcribing DNA into RNA  
 PT and translating RNA, or for direct in vitro RNA translation, where  
 PT protein components are labeled with a label pair adhering to each other.  
 XX  
 PS Disclosure; Page 44; 102pp; English.  
 XX  
 CC The invention describes a method of producing a peptide or its  
 CC derivative, comprising using a reaction system for transcribing DNA into  
 CC RNA and translating the RNA, or a reaction system for in vitro RNA  
 CC translation, where a part or all of protein components constituting the

CC reaction system are labeled with one of a pair of substances adhering to  
 CC each other and the other substance of the pair is used as an adsorbent  
 CC for capturing labeled protein components after translation. The produced  
 CC protein is efficiently isolated at a high purity from the reaction  
 CC system, and at the same time, the problem of consumption of energy in the  
 CC reaction system is resolved. This sequence represents a strep-tag which  
 CC binds to streptavidin and can be used to purify peptides in the peptide  
 CC production method of the invention  
 CC  
 SO Sequence 9 AA;  
 Query Match 88.6%; Score 39; DB 5; Length 9;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
 QY 2 XHPQFG 8  
 DB 3 RHPQFG 9  
 RESULT 8  
 AAG78476  
 ID AAG78476 standard; peptide; 9 AA.  
 XX  
 AC AAG78476;  
 XX  
 DT 18-JUN-2002 (first entry)  
 XX  
 DE Strep-tag nonapeptide linker sequence.  
 XX  
 KW Crystal lattice; crystallography; three dimensional structure;  
 KW membrane protein; PMB908.  
 XX  
 OS Synthetic.  
 XX  
 PN WO200185962-A1.  
 XX  
 PD 15-NOV-2001.  
 XX  
 PF 04-MAY-2001; 2001WO-GB002043.  
 XX  
 PR 05-MAY-2000; 2000SE-00001666.  
 XX  
 PR 02-JUN-2000; 2000US-0209331P.  
 XX  
 PR 28-JUN-2000; 2000SE-00002432.  
 XX  
 PA (IMCO-) IMPERIAL COLLEGE INNOVATIONS LTD.  
 XX  
 PA (IMAT/) IMAT S.  
 XX  
 PA (BYRN/) BYRNE B.  
 XX  
 PA (JORM/) JORMAKKA M.  
 XX  
 PA (ABRA/) ABRAMSON J.  
 XX  
 PA (SEJL/) SEJLITZ T.  
 XX  
 PI Iwata S, Byrne B, Jormakka M, Abramson J, Sejlitz T;  
 XX  
 DR WPI; 2002-089795/12.  
 XX  
 PT New recombinant vectors comprising promoter and nucleotide sequences,  
 PT useful in methods of crystallization, particularly for the  
 PT crystallization of proteins that are otherwise difficult to crystallize.  
 XX  
 PS Claim 19; Page 30; 70pp; English.  
 XX  
 CC This invention relates to recombinant vectors, comprising a promoter  
 CC sequence and a nucleotide sequence encoding a first protein, which is a  
 CC membrane protein, or multisubunit protein. The recombinant vector is  
 CC useful in methods of crystallization. The vector is particularly useful  
 CC for the crystallization of proteins that are otherwise difficult to  
 CC crystallize. This sequence represents the Strep-tag nonapeptide. This  
 CC linker acts as a bridge between subunit IV and the foreign fusion protein  
 CC of plasmid pMB908 (see AAH99982)  
 CC  
 SO Sequence 9 AA;

Query Match 88.6%; Score 39; DB 5; Length 9;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8  
 :|||||  
 DB 3 RHPQFG 9

## RESULT 9

AA016086 standard; peptide; 9 AA.

AC AA016086;

DT 27-FEB-2003 (first entry)

DE Neurological/CNS disease treatment method-related peptide #18.

XX Vaccine; gene therapy; neurological disease; CNS disorder;

KM central nervous system disorder; olfactory system; Alzheimer's disease;

KM Creutzfeldt-Jakob disease; Huntington's chorea; Parkinson's disease;

XX viral infection of the brain; brain tumour; lysosomal storage disease;

XX multiple sclerosis.

OS Unidentified.

PN WO200274243-A2.

PD 26-SEP-2002.

PF 15-MAR-2002; 2002WO-US008042.

PR 15-MAR-2001; 2001US-00808037.

XX (U9RA-) UNIV RAMOT APPLIED RES & IND DEV LTD.

PA (MCIN/) MCINNIS P.

PI Solomon B, Frenkel D;

XX WPI; 2003-040542/03.

XX The invention comprises a method for treating a neurological disease or a

CC central nervous system (CNS) disorder. The method involves displaying a

CC therapeutic molecule capable of treating the neurological disease or CNS

CC disorder on a viral display vehicle. The viral display vehicle is then

CC introduced into the olfactory system of a subject to treat the disease or

CC disorder. The method of the invention is useful for preventing, treating

CC and diagnosing neurological diseases or CNS disorders, such as:

CC Alzheimer's disease; Creutzfeldt-Jakob disease; Huntington's chorea; viral

CC infections of the brain; brain tumours; lysosomal storage diseases;

CC Parkinson's disease; and multiple sclerosis. The present amino acid

XX sequence represents a peptide which was used in the invention

XX SQ Sequence 9 AA;

## RESULT 10

ABP5547  
 ID ABP5547 standard; peptide; 9 AA.  
 XX  
 AC ABP5547;

DT 19-FEB-2003 (first entry)

DE Streptag peptide SEQ ID NO:77.

KM Hepatitis C virus; HCV; envelope protein; hepatotropic; virucide;

KM antiinflammatory; gene therapy; vaccine.

XX Hepatitis C virus.

OS Synthetic.

PN WO200285932-A2.

PD 31-OCT-2002.

PF 24-APR-2002; 2002WO-BE00062.

PR 24-APR-2001; 2001EP-00870088.

XX 17-JUL-2001; 2001US-0305604P.

XX (INNO-) INNOGENETICS NV.

PI Sablon E, Van Broekhoven A, Bosman A, Depla E, Deschamps G;

XX WPI; 2003-093095/08.

PT New recombinant nucleic acids for expressing Hepatitis C virus (HCV)

PT envelope proteins in eukaryotic cells, comprising a sequence encoding a

PT protein having an avian lysozyme leader peptide joined to the HCV

XX envelope protein.

XX Claim 5; Page 303; 319pp; English.

XX The present invention describes a recombinant nucleic acid (I) comprising

CC a nucleotide sequence encoding a protein having an avian lysozyme leader

CC peptide, or its functional equivalent, joined to a Hepatitis C virus

CC (HCV) envelope protein or its part. Also described: (1) a vector

CC comprising the recombinant nucleic acid; (2) a host cell comprising the

CC recombinant nucleic acid or the vector; and (3) a method for producing

CC HCV envelope protein or its part in a host cell, comprising transforming

CC the host cell with the recombinant nucleic acid or with the vector, where

CC the host cell is capable of expressing the protein cited above. (I) has

CC hepatotropic, virucide and antiinflammatory activities, and can be used

CC in gene therapy and vaccines. The recombinant nucleic acid is useful for

CC efficient expression of Hepatitis C virus envelope proteins in eukaryotic

CC cells, such as yeast cells. The HCV envelope proteins may be used as a

CC vaccine, or for incorporation into an immunoassay for the detection of

CC anti-HCV antibodies, and/or genotyping of HCV, for prognosing or

CC monitoring of HCV disease, or as a therapeutic agent. The method is used

CC for producing HCV envelope protein or its part in a host cell. AB084197

CC to AB084253 and ABP5528 to ABP5568 represent sequences used in the

XX exemplification of the present invention

XX SQ Sequence 9 AA;

Query Match 88.6%; Score 39; DB 6; Length 9;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFG 8  
 :|||||  
 DB 3 RHPQFG 9

## RESULT 11

AAE37229 standard; peptide; 9 AA.

XX AAE37229;

```

XX 07-AUG-2003 (first entry)
XX Strep-tag peptide.
DE Gene expression; therapy; isolation.
XX
XX Synthetic.
XX WO2003038049-A2.
XX 08-MAY-2003.
XX 29-OCT-2002; 2002WO-US034645.
XX 29-OCT-2001; 2001US-0340689P.
XX (RENO-) RENOVIS INC.
XX Heintz N, Serafini TA, Shyjan AW;
XX WPI; 2003-430512/40.
XX Isolating cell-type specific mRNA, useful in gene expression analysis or
XX quantification in a specific cell in a heterogeneous cell mixture, by
XX isolating mRNA bound to ribosomes or proteins that bind mRNA in a cell
XX type specific manner.
XX
XX Disclosure; Page 8; 136pp; English.
XX
XX The invention relates to a method for isolating mRNA from a population of
XX cells. The method involves selectively isolating ribosomes or proteins
XX that bind mRNA in a cell type specific manner and then isolating the mRNA
XX bound to the ribosomes or proteins that bind mRNA. The method is useful
XX for facilitating the analysis and quantification of gene expression in a
XX selected cell type present within a heterogeneous cell mixture. The
XX method may also be used in diagnostics or therapies for human diseases.
XX The present sequence is Strep-tag peptide. This sequence is used to
XX illustrate the method of the invention
XX
XX Sequence 9 AA;
SQ
Query Match 88.6%; Score 39; DB 6; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.8e+06; Mismatches 0; Gaps 0;
Matches 6; Conservative 1; Indels 0;
QY 2 XHPQFGG 8
Db 3 RHPQFGG 9

```

```

PR 17-JUL-2001; 2001US-0305604P.
XX (INNO-) INNOGENETICS NV.
XX Depia E, Bosman A, Deschamps G, Sablon E, Suckow M, Samson I;
XX Verheyden G;
XX WPI; 2003-103409/09.
XX
XX New Hepatitis C virus (HCV) envelope protein with N-glycosylation
XX site(s), useful as a vaccine for inducing a HCV-specific immune response
XX or HCV-specific antibodies, particularly for preventing or treating HCV
XX infection.
XX
XX Claim 10; Page 336; 355pp; English.
XX
XX The invention relates to Hepatitis C virus (HCV) envelope protein with N-
XX glycosylation site(s). The HCV envelope protein, or the pharmaceutical
XX composition comprising the envelope protein, is useful as a medicament or
XX a vaccine, particularly for inducing a HCV-specific immune response,
XX inducing HCV-specific antibodies or inducing a T-cell function in a
XX mammal. The protein is particularly useful for preventing, treating or
XX diagnosing HCV infection. It is also useful for detecting the presence of
XX anti-HCV antibodies in a sample. The present sequence is a peptide used
XX in the invention
XX
XX Sequence 9 AA;
SQ
Query Match 88.6%; Score 39; DB 6; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.8e+06; Mismatches 0; Gaps 0;
Matches 6; Conservative 1; Indels 0;
QY 2 XHPQFGG 8
Db 3 RHPQFGG 9

```

CC Agents that modulate activity of a membrane-spanning, signal-transducing  
CC (MST) protein. The method involves detecting a conformational change in  
CC a MST protein upon interaction with a ligand. The method is useful for  
CC identifying agents that modulate (e.g. agonists or antagonists) activity  
CC of MST protein. The present sequence is a strep epitope tag used to  
CC illustrate the method of the invention  
XX  
SQ Sequence 9 AA;  
  
Query Match 88.6%; Score 39; DB 6; Length 9;  
Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 XHPQFG 8  
DB 3 RHPQFG 9  
  
RESULT 14  
ABG72479  
ID ABG72479 standard; peptide; 9 AA.  
XX  
AC ABG72479;  
XX  
DT 18-FEB-2003 (first entry)  
XX  
DE Strep tag for use as a G protein epitope tag.  
XX  
XX G protein; alpha sub-unit; G protein coupled receptor; GPCR;  
KM G protein activation; G protein coupled receptor activation;  
KM G protein coupled receptor identification; strep tag.  
XX  
OS Synthetic.  
XX  
PN US6448377-B1.  
XX  
PD 10-SEP-2002.  
XX  
PF 27-SEP-2000; 2000US-00672239.  
XX  
PR 27-SEP-2000; 2000US-00672239.  
XX  
PA (STRD ) UNIV LEIAND STANFORD JUNIOR.  
XX  
PI Kobilka B, Lee TW,  
XX  
DR WPI; 2003-110149/10.  
XX  
PT Determining effects of candidate agent on activation of a G protein  
PT coupled receptor (GPCR) for evaluating new agonists and/or inverse  
PT agonists for GPCRs by contacting a candidate agent with a modified G  
PT protein alpha subunit and a GPCR.  
XX  
PS Disclosure; Col 9; 38pp; English.  
XX  
XX The invention describes a method of determining the effects of a  
CC candidate agent on activation of a G protein coupled receptor (GPCR). The  
CC method comprises contacting a candidate agent with a modified G protein  
CC alpha subunit and a GPCR, and detecting a level of G protein activation  
CC in response to the contacting, where the level of activation is  
CC indicative of the effects of the agent on the activity of GPCR. The  
CC method is useful for determining the effects of a candidate agent on  
CC activation of a G protein coupled receptor, evaluating new agonists,  
CC and/or inverse agonists for GPCRs, identifying ligands for GPCRs, and  
CC developing a strategy for identifying GPCRs involved in different  
CC biological processes, including diseases. The invention provides rapid  
CC and more sensitive bioassays for evaluating new agonists, agonists and/or  
CC inverse agonist for GPCRs. The method can be performed using membranes,  
CC which increases both the ease of performing the assay and its efficacy,  
CC and also allows high throughput screening of GPCR activity. Furthermore,  
CC this method directly measures GPCR activity, and thus is less labour-  
CC intensive than the conventional methods. This sequence represents the  
CC strep tag, an example of an epitope tag that can be used to tether the

CC stimulatory G protein alpha sub-unit to a membrane  
XX  
SQ Sequence 9 AA;  
  
Query Match 88.6%; Score 39; DB 6; Length 9;  
Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 XHPQFG 8  
DB 3 RHPQFG 9  
  
RESULT 15  
ADB84587  
ID ADB84587 standard; peptide; 9 AA.  
XX  
AC ADB84587;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Streptavidin conserved peptide #1.  
XX  
KM cell-free transcription system; cell-free translation system;  
KM protein synthesis; matrix; streptavidin.  
XX  
OS Escherichia coli.  
XX  
PN DE10137792-A1.  
XX  
PD 27-FEB-2003.  
XX  
PF 06-AUG-2001; 2001DE-01037792.  
XX  
PR 06-AUG-2001; 2001DE-01037792.  
XX  
PA (ERDM/) ERDMANN V.  
XX  
PI Erdmann VA, Lamla T, Stiege W,  
XX  
DR WPI; 2003-343999/33.  
XX  
PT Expressing genes in cell-free system, useful for preparation of proteins,  
PT comprises that the protein formed is removed from solution by binding to  
PT a matrix.  
XX  
XX Claim 13; Col 8; 8pp; German.  
XX  
XX This invention describes a novel method of expressing genes in a cell-  
CC free transcription and translation system which comprises using a  
CC reaction solution containing all necessary components of the  
CC transcription/translation system, amino acids, nucleotides and  
CC metabolites that supply energy and that are needed for synthesis. The  
CC proteins formed are immobilised on a matrix. The method allows simple  
CC recovery of proteins without a separate isolation step and the amount of  
CC protein produced can be determined before a reaction is complete.  
CC Continuous removal of proteins prevents it interfering with the  
CC expression process, making possible synthesis of proteins that interact  
CC adversely with the process, so normally produced only in very low yields.  
CC The use of a matrix concentrates the proteins produced and is applicable  
CC to proteins of any size. This sequence represents a highly conserved  
CC peptide from E. coli streptavidin which is used to illustrate the method  
CC of the invention.  
XX  
SQ Sequence 9 AA;  
  
Query Match 88.6%; Score 39; DB 7; Length 9;  
Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
QY 2 XHPQFG 8  
DB 3 RHPQFG 9

Wed Mar 2 14:27:15 2005

seq7.rag

Page 8

Search completed: March 2, 2005, 13:02:45  
Job time : 36.0325 secs

---



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 24.878 Seconds  
(without alignments)  
105.489 Million cell updates/sec

Title: SEQ7  
Perfect score: 44  
Sequence: 1 xhpqfgg 8

Scoring table: BLOSUM62X  
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*
- 2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*
- 6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09C\_NEW\_PUB.pep.\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*
- 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	88.6	8	14	US-10-026-578B-1
2	39	88.6	9	9	US-09-808-037-32
3	39	88.6	9	9	US-09-803-067-2
4	39	88.6	9	10	US-09-935-061-3
5	39	88.6	9	14	US-10-128-590-77
6	39	88.6	9	14	US-10-345-618-14
7	39	88.6	9	14	US-10-128-587A-77
8	39	88.6	9	15	US-10-384-788-32
9	39	88.6	9	15	US-10-449-831A-166
10	39	88.6	9	15	US-10-425-000-75
11	39	88.6	9	15	US-10-424-999-24
12	39	88.6	9	15	US-10-272-196-30
13	39	88.6	9	16	US-10-612-410-18

14	39	88.6	9	16	US-10-692-071-3	Sequence 3, Appl1
15	39	88.6	9	16	US-10-397-438A-6	Sequence 6, Appl1
16	39	88.6	9	17	US-10-753-309-5	Sequence 5, Appl1
17	39	88.6	9	17	US-10-494-248-5	Sequence 18, Appl1
18	39	88.6	10	9	US-09-801-968-18	Sequence 13, Appl1
19	39	88.6	10	10	US-09-802-154-18	Sequence 12, Appl1
20	39	88.6	10	13	US-10-060-765-13	Sequence 13, Appl1
21	39	88.6	10	14	US-10-263-230A-12	Sequence 13, Appl1
22	39	88.6	10	16	US-10-818-140-13	Sequence 13, Appl1
23	39	88.6	10	17	US-10-771-173-13	Sequence 8, Appl1
24	39	88.6	12	9	US-09-904-599A-8	Sequence 9, Appl1
25	39	88.6	18	16	US-10-397-438A-9	Sequence 1, Appl1
26	39	88.6	20	15	US-10-342-805-1	Sequence 2, Appl1
27	39	88.6	23	16	US-10-344-607-2	Sequence 16, Appl1
28	39	88.6	42	16	US-10-344-607-16	Sequence 20, Appl1
29	39	88.6	42	16	US-10-344-607-20	Sequence 12, Appl1
30	39	88.6	46	15	US-10-422-262-12	Sequence 8, Appl1
31	39	88.6	46	15	US-10-448-609-8	Sequence 57, Appl1
32	39	88.6	47	15	US-10-416-290-57	Sequence 212, App
33	39	88.6	132	15	US-10-449-831A-212	Sequence 11, Appl1
34	39	88.6	155	14	US-10-345-618-11	Sequence 13, Appl1
35	39	88.6	178	14	US-10-345-618-13	Sequence 13, Appl1
36	39	88.6	194	16	US-10-416-708A-10	Sequence 64, Appl1
37	39	88.6	194	16	US-10-416-708A-64	Sequence 8, Appl1
38	39	88.6	198	14	US-10-345-618-8	Sequence 24, Appl1
39	39	88.6	210	9	US-09-272-809-24	Sequence 198, App
40	39	88.6	212	15	US-10-449-831A-198	Sequence 192, App
41	39	88.6	234	15	US-10-449-831A-192	Sequence 204, App
42	39	88.6	329	15	US-10-449-831A-204	Sequence 4, Appl1
43	39	88.6	334	13	US-10-017-736-4	Sequence 4, Appl1
44	39	88.6	334	15	US-10-650-585-4	Sequence 6, Appl1
45	39	88.6	342	14	US-10-345-618-6	

## ALIGNMENTS

RESULT 1  
US-10-026-578B-1  
Sequence 1, Application US/10026578B  
Publication No. US20030083474A1  
GENERAL INFORMATION:  
APPLICANT: IBA (GmbH)  
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Tags  
FILE REFERENCE: 100810.01US1  
CURRENT APPLICATION NUMBER: US/10/026,578B  
CURRENT FILING DATE: 2002-11-11  
PRIOR APPLICATION NUMBER: DE 101 13 776.1  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: PCT/EP01/11846  
PRIOR FILING DATE: 2001-10-12  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 1  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Synthetic Peptide  
NAME/KEY: misc feature  
OTHER INFORMATION: Artificial Sequence represents peptide binding module  
US-10-026-578B-1

Query Match 88.6%; Score 39; DB 14; Length 8;  
Best Local Similarity 85.7%; Pred. No. 1.3e+06;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFGG 8  
:|||||  
DB 2 RHPQFGG 8

```
RESULT 2
US-09-808-037-32
/ Sequence 32, Application US/09808037
/ Patent No. US2002005231A1
/ GENERAL INFORMATION:
/ APPLICANT: SOLOMON, Beka
/ APPLICANT: HANAN, Elia
/ TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR THE TREATMENT AND/OR DIAGNOSIS OF
/ TITLE OF INVENTION: NEUROLOGICAL DISEASES AND DISORDERS
/ FILE REFERENCE: SOLOMON-2D
/ CURRENT APPLICATION NUMBER: US/09/808,037
/ CURRENT FILING DATE: 2001-03-15
/ PRIOR APPLICATION NUMBER: 09/629,971
/ PRIOR FILING DATE: 2000-07-31
/ PRIOR APPLICATION NUMBER: US 09/473,653
/ PRIOR FILING DATE: 1999-12-29
/ PRIOR APPLICATION NUMBER: US 60/152,417
/ PRIOR FILING DATE: 1999-09-03
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: Patehtin version 3.0
/ SEQ ID NO 32
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: synthetic construct
US-09-808-037-32

Query Match
Best Local Similarity 88.6%; Score 39; DB 9; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPQFG 8
Db 3 RHPQFG 9

RESULT 3
US-09-983-067-2
/ Sequence 2, Application US/09983067
/ Patent No. US2002012310A1
/ GENERAL INFORMATION:
/ APPLICANT: INOUE, Akio
/ APPLICANT: SHIMIZU, Yoshihiro
/ APPLICANT: UEDA, Takuya
/ TITLE OF INVENTION: Process for Producing Peptides By Using In Vitro
/ FILE REFERENCE: 1752-0151P
/ CURRENT APPLICATION NUMBER: US/09/983,067
/ CURRENT FILING DATE: 2001-10-23
/ PRIOR APPLICATION NUMBER: JP 294795/2001
/ PRIOR FILING DATE: 2001-09-26
/ PRIOR APPLICATION NUMBER: JP 227094/2001
/ PRIOR FILING DATE: 2001-07-27
/ PRIOR APPLICATION NUMBER: JP 6910/2001
/ PRIOR FILING DATE: 2001-01-15
/ PRIOR APPLICATION NUMBER: JP 401417/2000
/ PRIOR FILING DATE: 2000-12-28
/ NUMBER OF SEQ ID NOS: 6
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 2
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Strep-tag binding to streptavidin
/ NAME/KEY: MISC_FEATURE
/ LOCATION: (1)..(9)
/ OTHER INFORMATION: Schmidt & Skerra, 1993, "The random peptide library-assisted
/ OTHER INFORMATION: engineering of a C-terminal affinity peptide, useful for the
/ OTHER INFORMATION: detection and purification of a functional IgFv fragment",
US-09-983-067-2
```

```
Query Match
Best Local Similarity 88.6%; Score 39; DB 9; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPQFG 8
Db 3 RHPQFG 9

RESULT 4
US-09-935-061-3
/ Sequence 3, Application US/09935061
/ Publication No. US20030123649A1
/ GENERAL INFORMATION:
/ APPLICANT: Koblika, Brian M.
/ APPLICANT: Ghanouni, Pejman
/ APPLICANT: Lee, Tae Weon
/ TITLE OF INVENTION: Conformational assays to detect binding
/ TITLE OF INVENTION: to G protein-coupled receptors
/ FILE REFERENCE: STAN213
/ CURRENT APPLICATION NUMBER: US/09/935,061
/ CURRENT FILING DATE: 2001-08-21
/ PRIOR APPLICATION NUMBER: 60/286,250
/ PRIOR FILING DATE: 2001-04-24
/ NUMBER OF SEQ ID NOS: 16
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 3
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: epitope tag peptide
US-09-935-061-3

Query Match
Best Local Similarity 88.6%; Score 39; DB 10; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPQFG 8
Db 3 RHPQFG 9
```

```
RESULT 5
US-10-128-590-77
/ Sequence 77, Application US/10128590
/ Publication No. US20030108561A1
/ GENERAL INFORMATION:
/ APPLICANT: Immunogenetics N.V.
/ TITLE OF INVENTION: Core-glycosylated HCV envelope proteins
/ FILE REFERENCE: 135 PCT
/ CURRENT APPLICATION NUMBER: US/10/128,590
/ CURRENT FILING DATE: 2002-07-22
/ NUMBER OF SEQ ID NOS: 98
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 77
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: streptag
US-10-128-590-77

Query Match
Best Local Similarity 88.6%; Score 39; DB 14; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPQFG 8
Db 3 RHPQFG 9

RESULT 6
US-10-345-618-14
/ Sequence 14, Application US/10345618
```

```
/ Publication No. US2003014848A1
/ GENERAL INFORMATION:
/ APPLICANT: Koenigsen, Frank
/ APPLICANT: Sues, Gabriele M.
/ APPLICANT: Tarlinton, David M.
/ APPLICANT: Treutlein, Herbert R.
/ TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF PRODUCING SAME
/ FILE REFERENCE: 13474
/ CURRENT APPLICATION NUMBER: US/10/345,618
/ CURRENT FILING DATE: 2003-01-16
/ PRIOR APPLICATION NUMBER: US/09/509,031
/ PRIOR FILING DATE: 2000-06-09
/ NUMBER OF SEQ ID NOS: 16
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 14
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence:Strep-tag
US-10-345-618-14

Query Match      88.6%; Score 39; DB 14; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 7
US-10-128-587A-77
/ Sequence 77, Application US/10128587A
/ Publication No. US20030152940A1
/ GENERAL INFORMATION:
/ APPLICANT: Imogenece N.V.
/ TITLE OF INVENTION: Constructs and methods for expression of recombinant HCV envelope
/ FILE REFERENCE: 134 PCT
/ CURRENT APPLICATION NUMBER: US/10/128,587A
/ CURRENT FILING DATE: 2002-04-24
/ NUMBER OF SEQ ID NOS: 98
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 77
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: streptag
US-10-128-587A-77

Query Match      88.6%; Score 39; DB 14; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 8
US-10-384-788-32
/ Sequence 32, Application US/10384788
/ Publication No. US20040013647A1
/ GENERAL INFORMATION:
/ APPLICANT: SOLOMON, Bekk
/ APPLICANT: FRENKEL, Dan
/ TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR TREATING A PLAQUE-FORMING DISEASE
/ FILE REFERENCE: SOLOMON=2D.2
/ CURRENT APPLICATION NUMBER: US/10/384,788
/ CURRENT FILING DATE: 2003-03-11
/ PRIOR APPLICATION NUMBER: 60/371,735
```

```
/ PRIOR FILING DATE: 2002-04-12
/ PRIOR APPLICATION NUMBER: 09/808,037
/ PRIOR FILING DATE: 2001-03-15
/ PRIOR APPLICATION NUMBER: 09/830,954
/ PRIOR FILING DATE: 2001-06-22
/ PRIOR APPLICATION NUMBER: 10/162,889
/ PRIOR FILING DATE: 2002-06-06
/ PRIOR APPLICATION NUMBER: 09/473,653
/ PRIOR FILING DATE: 1998-12-29
/ PRIOR APPLICATION NUMBER: 09/629,971
/ PRIOR FILING DATE: 2000-07-31
/ PRIOR APPLICATION NUMBER: 60/152,417
/ PRIOR FILING DATE: 1999-09-03
/ PRIOR APPLICATION NUMBER: PCT/IL00/00518
/ PRIOR FILING DATE: 2000-08-31
/ NUMBER OF SEQ ID NOS: 33
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 32
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Synthetic construct
US-10-384-788-32

Query Match      88.6%; Score 39; DB 15; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 9
US-10-449-831A-166
/ Sequence 166, Application US/10449831A
/ Publication No. US20040029179A1
/ GENERAL INFORMATION:
/ APPLICANT: Koenigsen, Frank
/ TITLE OF INVENTION: Higher molecular weight entities and uses therefor
/ FILE REFERENCE: 2385978
/ CURRENT APPLICATION NUMBER: US/10/449,831A
/ CURRENT FILING DATE: 2003-05-30
/ PRIOR APPLICATION NUMBER: USSN 60/384878
/ PRIOR FILING DATE: 2002-05-31
/ NUMBER OF SEQ ID NOS: 237
/ SOFTWARE: PatentIn version 3.2
/ SEQ ID NO 166
/ LENGTH: 9
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Strep tag
US-10-449-831A-166

Query Match      88.6%; Score 39; DB 15; Length 9;
Best Local Similarity 85.7%; Pred. No. 1.3e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 10
US-10-425-000-75
/ Sequence 75, Application US/10425000
/ Publication No. US20040052777A1
/ GENERAL INFORMATION:
/ APPLICANT: Nebbit, Mark
/ APPLICANT: Cameron, Francis
/ APPLICANT: Blanche, Francis
/ TITLE OF INVENTION: Kringles Polypeptides and Methods for Using Them to Inhibit
/ TITLE OF INVENTION: Angiogenesis
```

```
FILE REFERENCE: ST01027-B
CURRENT APPLICATION NUMBER: US/10/425,000
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: 10/233,675
PRIOR FILING DATE: 2002-09-04
NUMBER OF SEQ ID NOS: 105
SOFTWARE: PatentIn version 3.2
SEQ ID NO 75
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Purification tag
US-10-425-000-75

Query Match
Best Local Similarity 88.6%; Score 39; DB 15; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
Db 3 RHPQFG 9

RESULT 11
US-10-424-999-24
Sequence 24, Application US/10/424,999
Publication No. US20040052810A1
GENERAL INFORMATION:
APPLICANT: Nesbitt, Mark
APPLICANT: Cameron, Beatrice
APPLICANT: Blanche, Francis
TITLE OF INVENTION: Abrogen Polypeptides, Nucleic Acids Encoding Them and Methods for
FILE REFERENCE: ST01027-A
CURRENT APPLICATION NUMBER: US/10/424,999
CURRENT FILING DATE: 2003-04-29
PRIOR APPLICATION NUMBER: 10/233,675
PRIOR FILING DATE: 2002-09-04
NUMBER OF SEQ ID NOS: 70
SOFTWARE: PatentIn version 3.2
SEQ ID NO 24
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Purification tag
US-10-424-999-24

Query Match
Best Local Similarity 88.6%; Score 39; DB 15; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
Db 3 RHPQFG 9

RESULT 12
US-10-272-196-30
Sequence 30, Application US/10/272,196
Publication No. US20040072746A1
GENERAL INFORMATION:
APPLICANT: Murray, Clare Margaret
APPLICANT: Hutchinson, Raymond
APPLICANT: Bantick, John Raymond
APPLICANT: Sullivan, Michael
APPLICANT: Donald, David Keith
APPLICANT: Jackson, Andrew Paul
APPLICANT: Jackson, Clive Geoffrey
APPLICANT: Cook, Ian David
TITLE OF INVENTION: INHIBITORS OF MONOCARBOXYLATE TRANSPORT
FILE REFERENCE: 06275-285001
```

```
CURRENT APPLICATION NUMBER: US/10/272,196
CURRENT FILING DATE: 2002-10-16
PRIOR APPLICATION NUMBER: US 60/329,318
PRIOR FILING DATE: 2001-10-16
NUMBER OF SEQ ID NOS: 44
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 30
LENGTH: 9
TYPE: PRT
ORGANISM: Homo sapiens
US-10-272-196-30

Query Match
Best Local Similarity 88.6%; Score 39; DB 15; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
Db 3 RHPQFG 9

RESULT 13
US-10-612-410-18
Sequence 18, Application US/10/612,410
Publication No. US20040132133A1
GENERAL INFORMATION:
APPLICANT: Bennett, Robert P.
TITLE OF INVENTION: Methods and Compositions for the Production, Identification and
FILE REFERENCE: 0942.5510003
CURRENT APPLICATION NUMBER: US/10/612,410
CURRENT FILING DATE: 2003-07-03
PRIOR APPLICATION NUMBER: 60/393,756
PRIOR FILING DATE: 2002-07-08
PRIOR APPLICATION NUMBER: 60/396,627
PRIOR FILING DATE: 2002-07-19
PRIOR APPLICATION NUMBER: 60/417,172
PRIOR FILING DATE: 2002-10-10
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn version 3.2
SEQ ID NO 18
LENGTH: 9
TYPE: PRT
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Strep epitope
US-10-612-410-18

Query Match
Best Local Similarity 88.6%; Score 39; DB 16; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
Db 3 RHPQFG 9

RESULT 14
US-10-692-071-3
Sequence 3, Application US/10/692,071
Publication No. US20040157268A1
GENERAL INFORMATION:
APPLICANT: Koblika, Brian K.
APPLICANT: Gnanou, Pejman
TITLE OF INVENTION: Conformational assays to detect binding
FILE REFERENCE: STAN-213CIP
CURRENT APPLICATION NUMBER: US/10/692,071
CURRENT FILING DATE: 2003-10-22
PRIOR APPLICATION NUMBER: PCT/US02/13250
PRIOR FILING DATE: 2002-04-24
PRIOR APPLICATION NUMBER: 09/935,061
```

; PRIOR FILING DATE: 2001-08-21  
 ; PRIOR APPLICATION NUMBER: 60/286,250  
 ; PRIOR FILING DATE: 2001-04-24  
 ; NUMBER OF SEQ ID NOS: 22  
 ; SOFTWARE: FastSeq for Windows Version 4.0  
 ; SEQ ID NO 3  
 ; LENGTH: 9  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: epitope tag peptide  
 US-10-692-071-3

Query Match 88.6%; Score 39; DB 16; Length 9;  
 Best Local Similarity 85.7%; Pred. No. 1.3e+06;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8  
 :|||||  
 DB 3 RHPQFG 9

RESULT 15  
 US-10-397-438A-6  
 ; Sequence 6, Application US/10397438A  
 ; Publication No. US20040191869A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Pharmacia & Upjohn AB  
 ; TITLE OF INVENTION: Fusion vectors  
 ; FILE REFERENCE: 00126  
 ; CURRENT APPLICATION NUMBER: US/10/397,438A  
 ; CURRENT FILING DATE: 2002-11-05  
 ; NUMBER OF SEQ ID NOS: 14  
 ; SOFTWARE: Patentin Ver. 2.1  
 ; SEQ ID NO 6  
 ; LENGTH: 9  
 ; TYPE: PRT  
 ; ORGANISM: Artificial Sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Artificial Sequence: Strep-tag  
 ; OTHER INFORMATION: linker  
 US-10-397-438A-6

Query Match 88.6%; Score 39; DB 16; Length 9;  
 Best Local Similarity 85.7%; Pred. No. 1.3e+06;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8  
 :|||||  
 DB 3 RHPQFG 9

Search completed: March 2, 2005, 14:18:51  
 Job time : 25.878 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 9.10569 Seconds

(without alignments) updates/sec  
65.585 Million cell updates/sec

Title: SEQ7  
Perfect score: 44  
Sequence: 1 rxhpqfsg 8

Scoring table:  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

## Database : Issued Patents AA:\*

1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/PTCUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	88.6	9	1	US-08-660-626-3
2	39	88.6	9	3	US-08-828-741B-14
3	39	88.6	9	3	US-08-948-097-1
4	39	88.6	9	3	US-09-031-168-3
5	39	88.6	9	3	US-09-160-567-14
6	39	88.6	9	4	US-09-672-235-3
7	39	88.6	9	4	US-09-710-299-14
8	39	88.6	9	4	US-09-509-031-14
9	39	88.6	9	4	US-09-669-516C-3
10	39	88.6	10	1	US-08-294-386C-11
11	39	88.6	10	2	US-08-737-316A-5
12	39	88.6	10	3	US-08-897-020-6
13	39	88.6	10	3	US-08-895-707-11
14	39	88.6	10	3	US-09-350-823-6
15	39	88.6	10	4	US-09-715-805-13
16	39	88.6	10	4	US-09-692-945-9
17	39	88.6	10	4	US-09-640-041-7
18	39	88.6	10	5	PCR-US95-10224-11
19	39	88.6	12	4	US-08-218-369-8
20	39	88.6	12	4	US-09-904-599A-8
21	39	88.6	12	5	PCR-US95-03742-8
22	39	88.6	15	1	US-08-664-448-33
23	39	88.6	15	3	US-08-828-741B-11
24	39	88.6	15	3	US-09-160-567-11
25	39	88.6	15	4	US-09-710-299-11
26	39	88.6	15	4	US-09-509-031-11
27	39	88.6	178	3	US-08-828-741B-13

28	39	88.6	178	3	US-09-160-567-13	Sequence 13, Appl
29	39	88.6	178	4	US-09-710-299-13	Sequence 13, Appl
30	39	88.6	178	4	US-09-509-031-13	Sequence 13, Appl
31	39	88.6	197	3	US-08-897-020-7	Sequence 7, Appl
32	39	88.6	197	3	US-09-350-823-7	Sequence 7, Appl
33	39	88.6	198	3	US-08-828-741B-8	Sequence 8, Appl
34	39	88.6	198	3	US-09-160-567-8	Sequence 8, Appl
35	39	88.6	198	4	US-09-710-299-8	Sequence 8, Appl
36	39	88.6	198	4	US-09-509-031-8	Sequence 8, Appl
37	39	88.6	334	4	US-10-017-736C-4	Sequence 6, Appl
38	39	88.6	342	3	US-08-828-741B-6	Sequence 6, Appl
39	39	88.6	342	3	US-09-160-567-6	Sequence 6, Appl
40	39	88.6	342	4	US-09-710-299-6	Sequence 6, Appl
41	39	88.6	342	4	US-09-509-031-6	Sequence 6, Appl
42	39	88.6	386	4	US-08-895-707-7	Sequence 7, Appl
43	39	88.6	409	4	US-10-017-736C-2	Sequence 2, Appl
44	39	88.6	487	4	US-09-270-767-43055	Sequence 43055, A
45	39	88.6	495	3	US-08-828-741B-4	Sequence 4, Appl

## ALIGNMENTS

RESULT 1  
US-08-660-626-3  
; Sequence 3, Application US/08660626  
; Patent No. 5789655  
; GENERAL INFORMATION:  
; APPLICANT: Stanley B. Prusiner  
; APPLICANT: Glenn C. Telling  
; APPLICANT: Fred E. Cohen  
; APPLICANT: Michael R. Scott  
; TITLE OF INVENTION: TRANSGENIC ANIMALS EXPRESSING  
; TITLE OF INVENTION: EPICTOP-TAGGED PROTEINS  
; NUMBER OF SEQUENCES: 13  
; CORRESPONDENCE ADDRESSES:  
; ADDRESSEE: Fish & Richardson  
; STREET: 2200 Sand Hill Road, Suite 100  
; CITY: Menlo Park  
; STATE: California  
; COUNTRY: USA  
; ZIP: 94025  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: ASCII  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/660,626  
; FILING DATE:  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Valeta Gregg  
; REGISTRATION NUMBER: 35, 127  
; REFERENCE/DOCKET NUMBER: 07532/003001  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (415) 322-5070  
; TELEFAX: (415) 854-0875  
; INFORMATION FOR SEQ ID NO: 3:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 9 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: peptide  
US-08-660-626-3  
Query Match  
Best Local Similarity 88.6%; Score 39; DB 1; Length 9;  
Matches 6; Conservative 1; Mismatches 0; Indels 0;  
Gaps 0;  
QY 2 XHPQFG 8  
:|||||  
Db 3 RHQFG 9

```

RESULT 2
US-08-828-741B-14
Sequence 14, Application US/08828741B
Patent No. 6043069
GENERAL INFORMATION:
APPLICANT: Koentgen, Frank
APPLICANT: Suenes, Gabriele M.
APPLICANT: Tariltein, David M.
APPLICANT: Treutlein, Herbert R.
TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
STREET: 400 Garden City Plaza
CITY: Garden City
STATE: New York
COUNTRY: United States of America
ZIP: 11530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/828,741B
FILING DATE: 26-MAR-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Digiglio, Frank S.
REGISTRATION NUMBER: 31,345
REFERENCE/DOCKET NUMBER: 10591
TELECOMMUNICATION INFORMATION:
TELEPHONE: (516) 742-4343
TELEFAX: (516) 742-4366
TELEX: 230 901 SANS UR
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 9 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-828-741B-14

Query Match 88.6%; Score 39; DB 3; Length 9;
Best Local Similarity 85.7%; Pred. No. 4, 1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0.

OY 2 XHPQFG 8
:|||||
Db 3 RHQFG 9

RESULT 3
US-08-948-097-1
Sequence 1, Application US/08948097C
Patent No. 6103493
GENERAL INFORMATION:
APPLICANT: Skerra, Arne
APPLICANT: Voss, Selma
TITLE OF INVENTION: Streptavidin Mutetins
FILE REFERENCE: HB8R 118
CURRENT APPLICATION NUMBER: US/08/948,097C
CURRENT FILING DATE: 1997-10-09
EARLIER APPLICATION NUMBER: DE 196 41 876.3
EARLIER FILING DATE: 1996-10-10
NUMBER OF SEQ ID NOS: 17
SEQ ID NO 1
LENGTH: 9
TYPE: PRT

```

```

ORGANISM: Artificial sequence
FEATURE:
; NAME/KEY: BINDING
; OTHER INFORMATION: Binding ligand for streptavidin
US-08-948-097-1

Query Match
Best Local Similarity 88.6%; Score 39; DB 3; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
:|||||
Db RHPQFG 9

RESULT 4
US-09-031-168-3
; Sequence 3, Application US/09031168
; Patent No. 6150583
; GENERAL INFORMATION:
; APPLICANT: Stanley B. Prusiner
; APPLICANT: Glenn C. Telling
; APPLICANT: Fred B. Cohen
; APPLICANT: Michael R. Scott
; TITLE OF INVENTION: "TRANSGENIC ANIMALS EXPRESSING
; EPTOPE-TAGGED PROTEINS
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 2200 Sand Hill Road, Suite 100
; CITY: Menlo Park
; STATE: California
; COUNTRY: USA
; ZIP: 94025
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Asciti
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/031.168
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/660,626
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Vaileta Gregg
; REGISTRATION NUMBER: 35,127
; REFERENCE/DOCKET NUMBER: 07532/003001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (415) 322-5070
; TELEFAX: (415) 854-0875
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 9 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; US-09-031-168-3

Query Match
Best Local Similarity 88.6%; Score 39; DB 3; Length 9;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFG 8
:|||||
Db RHPQFG 9

RESULT 5
US-09-160-567-14
; Sequence 14, Application US/09160567

```



```
Patent No. 6326179
GENERAL INFORMATION:
APPLICANT: Koenigen, Frank
APPLICANT: Sues, Gabriele M.
APPLICANT: Tarlinton, David M.
APPLICANT: Treutlein, Herbert R.
TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
  PRODUCING SAME
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
  ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
  STREET: 400 Garden City Plaza
  CITY: Garden City
  STATE: New York
  COUNTRY: United States of America
  ZIP: 11530
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/160,567
  FILING DATE:
  CLASSIFICATION:
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/828,741
  FILING DATE:
  ATTORNEY/AGENT INFORMATION:
  NAME: DIGIGLO, Frank S.
  REGISTRATION NUMBER: 31,346
  REFERENCE/DOCKET NUMBER: 10591
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: (516) 742-4343
  TELEFAX: (516) 742-4366
  TEXT: 230 901 SANS UR
  INFORMATION FOR SEQ ID NO: 14:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 9 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
US-09-160-567-14

Query Match      88.6%; Score 39; DB 3; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 6
US-09-672-239-3
Sequence 3, Application US/09672239
Patent No. 6448377
GENERAL INFORMATION:
APPLICANT: Lee, Tae Weon
APPLICANT: Koblik, Brian
TITLE OF INVENTION: MODIFIED G PROTEIN SUBUNITS
FILE REFERENCE: STAN-204
CURRENT APPLICATION NUMBER: US/09/672,239
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 9
TYPE: PRT
ORGANISM: Homo sapiens
US-09-672-239-3
```

```
Query Match      88.6%; Score 39; DB 4; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 7
US-09-710-299-14
Sequence 14, Application US/09710299
Patent No. 6521741
GENERAL INFORMATION:
APPLICANT: Koenigen, Frank
APPLICANT: Sues, Gabriele M.
APPLICANT: Tarlinton, David M.
APPLICANT: Treutlein, Herbert R.
TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
  PRODUCING SAME
NUMBER OF SEQUENCES: 14
CORRESPONDENCE ADDRESS:
  ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
  STREET: 400 Garden City Plaza
  CITY: Garden City
  STATE: New York
  COUNTRY: United States of America
  ZIP: 11530
COMPUTER READABLE FORM:
  MEDIUM TYPE: Floppy disk
  COMPUTER: IBM PC compatible
  OPERATING SYSTEM: PC-DOS/MS-DOS
  SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
  APPLICATION NUMBER: US/09/710,299
  FILING DATE: 09-No. 6521741-2000
  CLASSIFICATION: <Unknown>
  PRIOR APPLICATION DATA:
  APPLICATION NUMBER: 08/828,741
  FILING DATE: <Unknown>
  ATTORNEY/AGENT INFORMATION:
  NAME: DIGIGLO, Frank S.
  REGISTRATION NUMBER: 31,346
  REFERENCE/DOCKET NUMBER: 10591
  TELECOMMUNICATION INFORMATION:
  TELEPHONE: (516) 742-4343
  TELEFAX: (516) 742-4366
  TEXT: 230 901 SANS UR
  INFORMATION FOR SEQ ID NO: 14:
  SEQUENCE CHARACTERISTICS:
  LENGTH: 9 amino acids
  TYPE: amino acid
  STRANDEDNESS: single
  TOPOLOGY: linear
  MOLECULE TYPE: DNA (genomic)
  SEQUENCE DESCRIPTION: SEQ ID NO: 14:
US-09-710-299-14

Query Match      88.6%; Score 39; DB 4; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFG 8
       :|||||
Db      3 RHPQFG 9

RESULT 8
US-09-509-031-14
Sequence 14, Application US/09509031
Patent No. 6590080
GENERAL INFORMATION:
APPLICANT: Koenigen, Frank
```

```

; APPLICANT: Sueess, Gabriele M.
; APPLICANT: Tarlinton, David M.
; APPLICANT: Trentlehn, Herbert R.
; TITLE OF INVENTION: CATALYTIC ANTIBIOTIDS AND A METHOD OF PRODUCING SAME
; FILE REFERENCE: 13474
; CURRENT APPLICATION NUMBER: US/09/509,031
; CURRENT FILING DATE: 2000-06-09
; NUMBER OF SEQ ID NOS: 16
; SOFTWARE: Patentin Ver. 2.1
; SEQ ID NO 14
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Strep-tag
US-09-509-031-14

Query Match      88.6%; Score 39; DB 4; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFGG 8
       :|||||
DB      3 RHPQFGG 9

RESULT 9
US-09-669-516C-3
; Sequence 3, Application US/09669516C
; Patent No. 6603672
; GENERAL INFORMATION:
; APPLICANT: Prusiner, Stanley B.
; APPLICANT: Teiling, Glenn C.
; APPLICANT: Cohen, Fred E.
; APPLICANT: Scott, Michael R.
; TITLE OF INVENTION: RECOMBINANT CONSTRUCT ENCODING EPTOPE
; TITLE OF INVENTION: TAGGED PEP PROTEIN
; FILE REFERENCE: UOAL-045CON
; CURRENT APPLICATION NUMBER: US/09/669,516C
; CURRENT FILING DATE: 2000-09-25
; PRIOR APPLICATION NUMBER: 09/031,168
; PRIOR FILING DATE: 1998-02-26
; PRIOR APPLICATION NUMBER: 08/660,626
; PRIOR FILING DATE: 1996-06-06
; PRIOR APPLICATION NUMBER: 08/521,992
; PRIOR FILING DATE: 1995-08-31
; PRIOR APPLICATION NUMBER: 08/509,261
; PRIOR FILING DATE: 1995-07-31
; PRIOR APPLICATION NUMBER: 08/242,188
; PRIOR FILING DATE: 1994-05-13
; NUMBER OF SEQ ID NOS: 15
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 9
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Strep peptide
US-09-669-516C-3

Query Match      88.6%; Score 39; DB 4; Length 9;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFGG 8
       :|||||
DB      3 RHPQFGG 9

RESULT 10
US-08-294-386C-11
; Sequence 11, Application US/08294386C
; Patent No. 5646030
```

```

; GENERAL INFORMATION:
; APPLICANT: Ray, Bryan L.
; APPLICANT: Lin, Edmund C.C.
; APPLICANT: Crea, Roberto
; TITLE OF INVENTION: Method Of Isolating Mutant Cells
; NUMBER OF SEQUENCES: 20
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Iaplin & Kusmer
; STREET: 200 State Street
; CITY: Boston
; STATE: MA
; COUNTRY: U.S.A.
; ZIP: 02109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294,386C
; FILING DATE: August 23, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Kerner, Ann-Louise
; REGISTRATION NUMBER: 33,523
; REFERENCE/DOCKET NUMBER: SY22-010CIP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617/330-1300
; TELEFAX: 617/330-1311
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 10 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHEICAL: NO
US-08-294-386C-11

Query Match      88.6%; Score 39; DB 1; Length 10;
Best Local Similarity 85.7%; Pred. No. 0.66;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFGG 8
       :|||||
DB      4 RHPQFGG 10

RESULT 11
US-08-737-316A-5
; Sequence 5, Application US/08737316A
; Patent No. 5849576
; GENERAL INFORMATION:
; APPLICANT: SKERRA, Arne
; APPLICANT: WARDENBERG, Christina
; TITLE OF INVENTION: USE OF THE TETRACYCLINE PROMOTER FOR THE
; TITLE OF INVENTION: STRINGENTLY REGULATED PRODUCTION OF RECOMBINANT PROTEINS IN PR
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSER: Nikaido, Marmelstein, Murray & Oram LLP
; STREET: 655 Fifteenth Street N.W. Suite 330
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005-5701
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/737,316A
; FILING DATE: 12-NOV-1996
```

CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/EP95/01862  
FILING DATE: 17-MAY-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: DE P 44 17 598.1  
FILING DATE: 19-MAY-1994  
ATTORNEY/AGENT INFORMATION:  
NAME: Kiteb, Monica C.  
REGISTRATION NUMBER: 36,105  
REFERENCE/DOCKET NUMBER: P564-6012  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 638-5000  
TELEFAX: (202) 638-4810  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
US-08-737-316A-5

Query Match 88.6%; Score 39; DB 2; Length 10;  
Best Local Similarity 85.7%; Pred. No. 0.66;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPORFG 8  
DB 4 RHPORFG 10

RESULT 12  
US-08-897-020-6  
Sequence 6, Application US/08897020  
Patent No. 6028176  
GENERAL INFORMATION:  
APPLICANT: Shanafelt, Armen; Greve, Jeffrey; Rocznik, Steven  
TITLE OF INVENTION: High-affinity Interleukin-4 Mutains  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Bayer Corporation, Pharmaceutical Division  
STREET: 400 Morgan Lane  
CITY: West Haven  
STATE: CT  
COUNTRY: United States of America  
ZIP: 06516-4175  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS V. 6.30  
SOFTWARE: Word for Windows 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/897,020  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: P-91,242  
FILING DATE: 19-JUL-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Huw R. Jones  
REGISTRATION NUMBER: 33, 916  
REFERENCE/DOCKET NUMBER: WH5020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (203) 812-2317  
TELEFAX: (203) 812-5492  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
DESCRIPTION: tag for streptavidin  
HYPOTHETICAL: no

ANTI-SENSE: no  
US-08-897-020-6

Query Match 88.6%; Score 39; DB 3; Length 10;  
Best Local Similarity 85.7%; Pred. No. 0.66;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPORFG 8  
DB 4 RHPORFG 10

RESULT 13  
US-08-895-707-11  
Sequence 11, Application US/08895707  
Patent No. 6077700  
GENERAL INFORMATION:  
APPLICANT: (Pharmacia & Upjohn, Co.)  
APPLICANT: alternatively, for U.S. filing:  
APPLICANT: Hollingsworth, Robert A.  
APPLICANT: Sharma, Satish K.  
APPLICANT: Rank, Kenneth B.  
APPLICANT: Evans, David B.  
TITLE OF INVENTION: Special Constructs and Complexes of  
NUMBER OF SEQUENCES: 21  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Pharmacia & Upjohn Company  
STREET: 301 Henrietta Street  
CITY: Kalamazoo  
STATE: MI  
COUNTRY: USA  
ZIP: 49001  
COMPUTER READABLE FORM:  
MEDIUM TYPE: floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Rel. #1.0, Ver. #1.25/WordPerfect 5.2+  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/895,707  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Woonton, Thomas A.  
REGISTRATION NUMBER: 35,004  
REFERENCE/DOCKET NUMBER: 6054  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 616-833-7914  
TELEFAX: 616-833-8897  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
US-08-895-707-11

Query Match 88.6%; Score 39; DB 3; Length 10;  
Best Local Similarity 85.7%; Pred. No. 0.66;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPORFG 8  
DB 4 RHPORFG 10

RESULT 14  
US-09-350-823-6  
Sequence 6, Application US/09350823

Patent No. 6313272  
GENERAL INFORMATION:  
APPLICANT: Shantafelt, Armen; Greve, Jeffrey; Rocznak, Steven  
TITLE OF INVENTION: High-affinity Interleukin-4 Muteins  
NUMBER OF SEQUENCES: 9  
CORRESPONDENCE ADDRESSES:  
ADDRESSEE: Bayer Corporation, Pharmaceutical Division  
STREET: 400 Morgan Lane  
CITY: West Haven  
STATE: CT  
COUNTRY: United States of America  
ZIP: 06516-4175  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette, 3.5 inch, 1.44 Mb storage  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS V. 6.30  
SOFTWARE: Word for Windows 6.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/09/350,823  
FILING DATE: 09-Jul-1999  
CLASSIFICATION: <Unknown>  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: 08/897,020  
FILING DATE: <Unknown>  
ATTORNEY/AGENT INFORMATION:  
NAME: Huw R. Jones  
REGISTRATION NUMBER: 33, 916  
REFERENCE/DOCKET NUMBER: WH5020  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (203) 812-2317  
TELEFAX: (203) 812-5492  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 10  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
DESCRIPTION: tag for streptavidin  
HYPOTHETICAL: no  
ANTI-SENSE: no  
SEQUENCE DESCRIPTION: SEQ ID NO: 6:  
US-09-350-823-6  
Query Match 88.6%; Score 39; DB 3; Length 10;  
Best Local Similarity 85.7%; Pred. No. 0.66;  
Matches 6; Conservative 1; Mismatches 0; Gaps 0;  
QY 2 XHPQFGG 8  
:|||||  
Db 4 RHPQFGG 10  
RESULT 15  
US-09-715-805-13  
Sequence 13, Application US/09715805  
Patent No. 6716626  
GENERAL INFORMATION:  
APPLICANT: Itoh, No. 6716626uyuki  
APPLICANT: Kavanagh, W. Michael  
TITLE OF INVENTION: HUMAN FGF-21 GENE AND GENE EXPRESSION  
TITLE OF INVENTION: PRODUCTS  
FILE REFERENCES: PP-16758,001/201130,408  
CURRENT APPLICATION NUMBER: US/09/715,805  
NUMBER OF SEQ ID NOS: 17  
CURRENT FILING DATE: 2000-11-16  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 13  
LENGTH: 10  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Residues which bind to paramagnetic streptavidin

OTHER INFORMATION: beads (used for purification).  
US-09-715-805-13  
Query Match 88.6%; Score 39; DB 4; Length 10;  
Best Local Similarity 85.7%; Pred. No. 0.66;  
Matches 6; Conservative 1; Mismatches 0; Gaps 0;  
QY 2 XHPQFGG 8  
:|||||  
Db 4 RHPQFGG 10  
Search completed: March 2, 2005, 12:25:34  
Job time: 9.10569 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 6.11382 Seconds  
(without alignments)  
125.901 Million cell updates/sec

Title: SEQ8

Perfect score: 42

Sequence: 1 rxhpgfek 8

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:\*

1: p1r1:\*

2: p1r2:\*

3: p1r3:\*

4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	90.5	300	2	G75436 conserved hypothet
2	36	85.7	332	2	AE3304 oxidoreductase (EC
3	36	85.7	360	2	B71073 probable malate de
4	35	85.7	529	2	T14947 hypothetical prote
5	35	83.3	129	2	A82547 hypothetical prote
6	34	81.0	189	2	T36642 hypothetical prote
7	34	81.0	213	2	AD2533 hypothetical prote
8	34	81.0	362	2	H75131 malate dehydrogena
9	34	81.0	437	2	A31142 gelatin, ovarian
10	34	81.0	444	2	T10487 glutamate dehydrog
11	34	81.0	543	2	A84134 acetolactate synth
12	34	81.0	627	2	T00484 hypothetical prote
13	34	81.0	1473	2	T38791 probable ferredoxi
14	34	81.0	3746	1	YGPLV3 alpha-aminoacylpyl-
15	34	81.0	3791	1	YGPLV8 alpha-aminoacylpyl-
16	33	78.6	101	2	T10856 carboxypeptidase C
17	33	78.6	211	2	C84162 hypothetical prote
18	33	78.6	310	2	C84701 hypothetical prote
19	33	78.6	335	2	AC3160 NAD binding oxidor
20	33	78.6	347	2	B85640 hypothetical prote
21	33	78.6	452	2	F84421 hypothetical prote
22	33	78.6	875	2	T50182 ubiquitin-specific
23	33	78.6	2139	2	T18296 myosin heavy chain
24	32	76.2	69	2	C82724 hypothetical prote
25	32	76.2	115	2	T15519 hypothetical prote
26	32	76.2	159	2	D75278 hypothetical prote
27	32	76.2	203	2	JQ1973 HTP9-A protein - m
28	32	76.2	293	2	E82582 DnaJ protein XF23
29	32	76.2	312	2	S67052 hypothetical prote

30	32	76.2	313	2	T05139 hypothetical prote
31	32	76.2	324	2	T08307 hypothetical prote
32	32	76.2	338	2	T30538 heat shock protein
33	32	76.2	346	2	S09539 DNA ligase (ATP) (
34	32	76.2	354	2	A86843 prephenate dehydro
35	32	76.2	356	2	T38408 hypothetical prote
36	32	76.2	368	2	S75132 sensory transducti
37	32	76.2	385	2	A44102 di-n-acetylchitobi
38	32	76.2	406	2	AB1060 arginine deiminase
39	32	76.2	420	2	S71199 dnaJ protein homol
40	32	76.2	434	2	H81709 N utilization subs
41	32	76.2	434	2	A71559 probable transcrip
42	32	76.2	441	2	C75076 heme biosynthesis
43	32	76.2	481	2	T14300 hypothetical prote
44	32	76.2	486	1	A35667 ty transcriptional p
45	32	76.2	491	2	T34499 hypothetical prote

#### ALIGNMENTS

```

RESULT 1
G75436
conserved hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Jul-2004
C:Accession: G75436
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
, M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Usterback, T.; Zaleski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.
A:Reference number: AF250; MUID:20036896; PMID:10567266
A:Accession: G75436
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-300 <WHI>
A:References: UNIPROT:Q9RVC9; GB:AE001960; GB:AE000513; NID:G6458833; PIDN:AAF1067
A:Experimental source: strain R1
C:Genetics:
A:Gene: DR1100
A:Map position: 1
C:Superfamily: Escherichia coli ycea protein

Query Match      90.5%; Score 38; DB 2; Length 300;
Best Local Similarity 75.0%; Pred. No. 8;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 RXHPQFEK 8
DB      261 RAHPQFEK 268

RESULT 2
AE3304
oxidoreductase (EC 1.1.1.-) [imported] - Brucella melitensis (strain 16M)
C:Species: Brucella melitensis
C:Date: 01-Feb-2002 #sequence_revision 01-Feb-2002 #text_change 09-Jul-2004
C:Accession: AF3304
R:DelVecchio, V.G.; Kapral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.; Ivanova,
, Marur, M.; Goltzman, E.; Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letess
Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002
A>Title: The genome sequence of the facultative intracellular pathogen Brucella melitensis
A:Reference number: AD3252; PMID:11756688
A:Accession: AF3304
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-332 <KUR>
A:Cross-References: UNIPROT:Q8Y1M3; GB:AB008917; PIDN:AAL51601.1; PID:gl7982326; GSPDB:(
A:Experimental source: strain 16M
C:Genetics:
A:Gene: BMEI0420
A:Map position: 1

```

C:Keywords: oxidoreductase

Query Match 85.7%; Score 36; DB 2; Length 332;  
Best Local Similarity 75.0%; Pred. No. 22;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPQFEK 8  
DB 135 RHPAFAFEK 142

RESULT 3

B71073  
probable malate dehydrogenase - *Pyrococcus horikoshii*

C:Species: *Pyrococcus horikoshii*  
C:Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 12-Jul-2004

C:Accession: B71073

R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hatakeyama, Y.; Hino, Y.; Yamamoto, S.; Sekita, M.; Ohnuki, Y.; Funahashi, T.; Tanaka, T.; Kudo, Y.; Yamazaki, J.; Kushida, N.; Oguchi, M.; DNA Res. 5, 55-76, 1998

A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic *Pyrococcus* strain

A:Reference number: A71000; MUID:98344137; PMID:9679194

A:Accession: B71073

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-360 <KAN>

A:Cross-references: UNIPROT:O59028; GB:AP000005; NID:G3236132; PIDN:BAA30380.1; PID:G3236132

A:Experimental source: strain OT3

A>Note: this accession replaces an interim accession for a sequence replaced by GenBank

C:Genetics:

A:Gene: PH1277

Query Match 85.7%; Score 36; DB 2; Length 360;  
Best Local Similarity 62.5%; Pred. No. 24;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPQFEK 8  
DB 311 RHPAFAFEK 318

RESULT 4

T14947  
hypothetical protein Y1031 - *Yersinia pestis* plasmid pMT1

C:Species: *Yersinia pestis*  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004

C:Accession: T14947

R:Lindegger, L.E.; Plano, G.V.; Burland, V.; Mayhew, G.F.; Blattner, F.R.

A:Title: Complete DNA sequence and detailed analysis of the *Yersinia pestis* KIMS plasmid

A:Reference number: Z18268; MUID:99043898; PMID:9826348

A:Accession: T14947

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-529 <LIN>

A:Cross-references: UNIPROT:O68740; EMBL:AF074611; NID:G3883003; PID:G3883032; PIDN:AAAC

C:Genetics:

A:Gene: Y1031

A:Genome: plasmid pMT1

Query Match 85.7%; Score 36; DB 2; Length 529;  
Best Local Similarity 62.5%; Pred. No. 36;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPQFEK 8  
DB 19 RHPAFAFEK 26

RESULT 5

A82547  
hypothetical protein XF2515 [imported] - *Xylella fastidiosa* (strain 9ASC)  
C:Species: *Xylella fastidiosa*

C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004

C:Accession: A82547

R:anonymous, The *Xylella fastidiosa* Consortium of the Organization for Nucleotide Sequencing

A:Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A>Note: for a complete list of authors see reference number A59328 below

A:Accession: A82547

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-129 <SIM>

A:Cross-references: UNIPROT:Q9PAX0; GB:AE004059; GB:AE003849; NID:G9107718; PIDN:AAF8531

A:Experimental source: strain 9ASC

R:Simpson, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvares, R.; A. Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Canario, L.E.A.; Carraro, D.M.; Carre, H.

as-Veto, E.; Docena, C.; El-Dorri, H.; Facinelli, A.P.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franco, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Krieger, J.E.; Kuramae, E.B.; Laig

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E

A:Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.

, F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.J. de M.; de Rosa, V.E.; de Sa, R.G.; Santelli, R.V.; Sawasak

M.; Tsuchioka, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF2515

Query Match 83.3%; Score 35; DB 2; Length 129;  
Best Local Similarity 62.5%; Pred. No. 12;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPQFEK 8  
DB 94 RHPAFAFEK 101

RESULT 6

T36642  
hypothetical protein SCH35.17 - *Streptomyces coelicolor*

C:Species: *Streptomyces coelicolor*  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004

C:Accession: T36642

R:Oliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrett, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, June 1999

A:Accession: T36642

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-189 <OLI>

A:Cross-references: UNIPROT:Q9X8Y9; EMBL:AL078610; PIDN:CAB44410.1; GSPDB:GN00070; SCOPED

A:Experimental source: strain A3(2)

C:Genetics:

A:Gene: SCODEB:SCH35.17

Query Match 81.0%; Score 34; DB 2; Length 189;  
Best Local Similarity 71.4%; Pred. No. 29;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPQFEK 7  
DB 105 RHPAFAFEK 111

RESULT 7

AD2533  
hypothetical protein alr7551 [imported] - *Nostoc* sp. (strain PCC 7120) plasmid pCC7120be

C:Species: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

A>Note: Nostoc sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004

C:Accession: AD2533

R.Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Matanabe, A.; Irituguhi, Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001

A1Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Ana*

A:Accession: AB1807; MUID:21595285; PMID:11759840

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-213 <KUR>

A:Cross-references: UNIPROT:Q8ZSF9; GB:AP003602; PIDN:BAW77194.1; PID:G17134636; GSPDB:G

C:Genetics:

A:Gene: alx7551

A:Genome: plasmid

Query Match 81.0%; Score 34; DB 2; Length 213;  
Best Local Similarity 71.4%; Pred. No. 33;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 1 RXHPOFE 7  
:|||||:  
Db 114 RWHPOFO 120

RESULT 8

H75131

malate dehydrogenase PAB1791 - *Pyrococcus abyssi* (strain Oraz)

C:Species: *Pyrococcus abyssi*

C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004

C:Accession: H75131

R:Anonymous, Gendecore submitted to the EMBL Data Library, July 1999

A:Description: *Pyrococcus abyssi* genome sequence: insights into archaeal chromosome stru

A:Reference number: A75001

A:Accession: H75131

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-362 <KAW>

A:Cross-references: UNIPROT:Q9V0D5; GB:AJ248285; GB:AL096836; NID:G5458067; PIDN:CAW4976

A:Experimental source: strain Oraz

C:Genetics:

A:Gene: PAB1791

C:Superfamily: malate dehydrogenase ylbC

Query Match 81.0%; Score 34; DB 2; Length 362;  
Best Local Similarity 62.5%; Pred. No. 59;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

OY 1 RXHPOFE 8  
:|||||:  
Db 311 RHPDPER 318

RESULT 9

AJ1142

gelsolin - African clawed frog (fragments)

C:Species: *Xenopus laevis* (African clawed frog)

C>Date: 31-Mar-1990 #sequence\_revision 31-Mar-1990 #text\_change 09-Jul-2004

C:Accession: AJ1142; BJ1142

R:Ankenbauer, T.; Kleinschmidt, J.A.; Vandekerckhove, J.; Franke, W.W.

J. Cell Biol. 107, 1489-1498, 1988

A1Title: Proteins regulating actin assembly in oogenesis and early embryogenesis of *Xen*

A:Reference number: AJ1142; MUID:89008590; PMID:2844829

A:Accession: AJ1142

A:Molecule type: mRNA

A:Residues: 21-437 <ANK1>

A:Cross-references: UNIPROT:P14885; EMBL:X13319; NID:G64715; PIDN:CAA31694.1; PID:G64716

A:Accession: BJ1142

A:Molecule type: protein

A:Residues: 1-20 <ANK2>

C:Superfamily: gelsolin; gelsolin repeat homology

C:Keywords: actin binding

P:21-53/Domain: gelsolin repeat homology (fragment) <GEL1>

F:101-419/Domain: gelsolin repeat homology <GEL2>

Query Match 81.0%; Score 34; DB 2; Length 437;  
Best Local Similarity 71.4%; Pred. No. 72;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPOFEK 8  
:|||||:  
Db 3 DHPDPER 9

RESULT 10

T10487

glutamate dehydrogenase (NADP) (EC 1.4.1.4) - *Prevotella ruminicola*

C:Species: *Prevotella ruminicola*

C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004

C:Accession: T10487

R:Wen, Z.T.; Morrison, M. submitted to the EMBL Data Library, December 1996

A:Reference number: Z17049

A:Accession: T10487

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-444 <WEN>

A:Cross-references: UNIPROT:P95544; EMBL:U82240; NID:G1772844; PID:G1772845

A:Experimental source: strain B14

C:Genetics:

A:Note: gdhA

C:Superfamily: glutamate dehydrogenase (NAD(P)+)

C:Keywords: NADP; oxidoreductase

Query Match 81.0%; Score 34; DB 2; Length 444;  
Best Local Similarity 71.4%; Pred. No. 73;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPOFEK 8  
:|||||:  
Db 37 KHPEFEK 43

RESULT 11

A84134

acetylacetate synthase large subunit BH3873 [imported] - *Bacillus halodurans* (strain C-12)

C:Species: *Bacillus halodurans*

C>Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 16-Aug-2004

C:Accession: A84134

R:Takami, H.; Nakasone, K.; Takaki, Y.; Meeno, G.; Sasaki, R.; Maui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A1Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: A84134

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-543 <STO>

A:Cross-references: UNIPROT:Q9K659; GB:AP001520; GB:BA000004; NID:G10176401; PIDN:BAW075

C:Genetics:

A:Experimental source: strain C-125

C:Superfamily: Acetylacetate synthase, large subunit/pyruvate oxidase

Query Match 81.0%; Score 34; DB 2; Length 543;  
Best Local Similarity 62.5%; Pred. No. 91;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

OY 1 RXHPOFEK 8  
:|||||:  
Db 470 RHPDPER 477

RESULT 12

T00484

hypothetical protein At2G5030 [imported] - *Arabidopsis thaliana*

N:Alternate names: hypothetical protein F1913.26

C.Species: Arabidopsis thaliana (mouse-ear cress)  
 C.Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004  
 C.Accession: T00484; G84763  
 R.Roundley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul  
 submitted to the EMBL Data Library, April 1998  
 A.Description: Arabidopsis thaliana chromosome II BAC F1913 genomic sequence.  
 A.Reference number: Z14160  
 A.Accession: T00484  
 A.Status: translated from GB/EMBL/DBJ  
 A.Molecule type: DNA  
 A.Residues: 1-627 <ROU>  
 A.Cross-references: UNIPROT:064766; EMBL:AC004238; NID:g3033373; PID:g3033399  
 A.Experimental source: cultivar Columbia  
 R.Lin, X.; Kaul, S.; Roundley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.;  
 M.; Koo, H.; Mofatt, K.S.; Cronin, L.A.; Shen, M.; Vanden, S.E.; Umayam, L.; Talion, L.  
 eus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.  
 Nature 402, 761-766, 1999  
 A.Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.  
 A.Reference number: A84420; MUID:20083487; PMID:10617197  
 A.Accession: G84763  
 A.Status: preliminary  
 A.Molecule type: DNA  
 A.Residues: 1-627 <STO>  
 A.Cross-references: GB:AE002093; NID:g3033399; PID:AA12843.1; GSPDB:GN00139  
 C.Genetics:  
 A.Gene: F1913.26; Atg530503  
 A.Map position: 2

Query Match  
 Best Local Similarity 81.0%; Score 34; DB 2; Length 627;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFE 7  
 DB 590 KSHPOFE 596

RESULT 13  
 T38791  
 Probable ferredoxin oxidoreductase SPAC4C5.05c - fission yeast (Schizosaccharomyces pombe)  
 C.Species: Schizosaccharomyces pombe  
 C.Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C.Accession: T38791; T37495  
 R.Pearson, D.; Churcher, C.M.; Barrell, B.G.; Rajandream Wood, V.  
 submitted to the EMBL Data Library, August 1997  
 A.Reference number: Z21811  
 A.Accession: T38791  
 A.Status: preliminary; translated from GB/EMBL/DBJ  
 A.Molecule type: DNA  
 A.Residues: 1-1473 <PEA>  
 A.Cross-references: UNIPROT:O14167; EMBL:Z98560; PID:CA11176.1; GSPDB:GN00066; SPDB:SF  
 A.Experimental source: strain 972h-; cosmid c4C5  
 R.Connor, R.; Churcher, C.M.; Barrell, B.G.; Rajandream, M.A.; Wood, V.  
 submitted to the EMBL Data Library, November 1997  
 A.Reference number: Z21719  
 A.Accession: T37495  
 A.Status: preliminary  
 A.Molecule type: DNA  
 A.Residues: 1-755 <CON>  
 A.Cross-references: EMBL:AL009197; PID:CAA15714.1; GSPDB:GN00066; SPDB:SPAC10F6.01c  
 C.Genetics:  
 A.Gene: SPDB:SPAC4C5.05c  
 A.Map position: 1

Query Match  
 Best Local Similarity 81.0%; Score 34; DB 2; Length 1473;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFE 8  
 DB 1221 RDHPOFE 1228

RESULT 14  
 YGPLY3  
 alpha-aminoadipyl-cysteine-lysin synthetase (EC 6.-.-.-) - Penicillium chrysogenum (str  
 N/Alternate names: ACV synthetase.  
 C.Species: Penicillium chrysogenum  
 C.Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 12-Jul-2004  
 C.Accession: S13134  
 R.Smith, D.J.; Earl, A.J.; Turner, G.  
 EMBO J. 9, 2743-2750, 1990  
 A>Title: The multifunctional peptide synthetase performing the first step of penicillin  
 ic synthetases.  
 A.Reference number: S13134; MUID:90360984; PMID:2118102  
 A.Accession: S13134  
 A.Molecule type: DNA  
 A.Residues: 1-3746 <SMI>  
 A.Cross-references: UNIPROT:P19787; GB:X54296; NID:g3117; PID:CAA38195.1; PID:g3118  
 C.Comment: This multifunctional enzyme catalyzes the formation of alpha-aminoadipyl-cyst  
 e-tying out the polymerization steps to form the tripeptide, which is the first common int  
 C.Genetics:  
 A.Gene: pcbAB  
 C.Superfamily: non-ribosomal peptide synthetase; acetate-CoA ligase homology; acyl carri  
 C.Keywords: carrier protein; cephalosporin biosynthesis; duplication; ligase; penicillin  
 F1338-804/Domain: acetate-CoA ligase homology <ACLI>  
 F1821-891/Domain: acyl carrier protein homology <ACPI>  
 F1432-1889/Domain: acetate-CoA ligase homology <ACLI2>  
 F1905-1975/Domain: acyl carrier protein homology <ACPI2>  
 F12515-2977/Domain: acetate-CoA ligase homology <ACLI3>  
 F12994-3062/Domain: acyl carrier protein homology <ACPI3>  
 F1855,1939,3026/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match  
 Best Local Similarity 81.0%; Score 34; DB 1; Length 3746;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFE 7  
 DB 2272 RTHPOFE 2278

RESULT 15  
 YGPLY8  
 alpha-aminoadipyl-cysteine-lysin synthetase (EC 6.-.-.-) - Penicillium chrysogenum (str  
 N/Alternate names: ACV synthetase  
 C.Species: Penicillium chrysogenum  
 C.Date: 31-Dec-1991 #sequence\_revision 31-Dec-1991 #text\_change 12-Jul-2004  
 C.Accession: A37886  
 R.Diez, B.; Gutierrez, S.; Barredo, J.L.; van Solingen, P.; van der Voort, L.H.M.; Marti  
 J. Biol. Chem. 265, 16358-16365, 1990  
 A>Title: The cluster of penicillin biosynthetic genes. Identification and characterizati  
 on genes.  
 A.Reference number: A37886; MUID:90375501; PMID:2129535  
 A.Accession: A37886  
 A.Molecule type: DNA  
 A.Residues: 1-3791 <DIE>  
 A.Cross-references: UNIPROT:P26046; GB:M57425; GB:U05604; NID:g169183; PID:AAA63415.1;  
 C.Comment: This multifunctional enzyme catalyzes the formation of alpha-aminoadipyl-cyst  
 e-tying out the polymerization steps to form the tripeptide, which is the first common int  
 C.Genetics:  
 A.Gene: pcbAB  
 C.Superfamily: non-ribosomal peptide synthetase; acetate-CoA ligase homology; acyl carri  
 C.Keywords: carrier protein; cephalosporin biosynthesis; duplication; ligase; penicillin  
 F1368-834/Domain: acetate-CoA ligase homology <ACLI>  
 F1851-923/Domain: acyl carrier protein homology <ACPI>  
 F1464-1921/Domain: acetate-CoA ligase homology <ACLI2>  
 F1937-2007/Domain: acyl carrier protein homology <ACPI2>  
 F12547-3009/Domain: acetate-CoA ligase homology <ACLI3>  
 F13026-3094/Domain: acyl carrier protein homology <ACPI3>  
 F1885,1971,3058/Binding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match  
 Best Local Similarity 81.0%; Score 34; DB 1; Length 3791;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;



Oy 1 RXHPOFE 7  
|:|:|:|:  
Db 2304 RTHPOFO 2310

Search completed: March 2, 2005, 12:28:54  
Job time : 7.11382 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 29.7236 Seconds

(without alignments)  
137.824 Million cell updates/sec

Title: SEQ8

Perfect score: 42

Sequence: 1 rxhpgfek 8

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trcemb:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	190.5	300	1 YB00 DEIRA	09rvy3 deinococcus
2	37	88.1	704	2 08CJD2	08cjd2 rattus norv
3	37	88.1	743	2 08CH90	08ch90 rattus norv
4	36	88.1	2058	2 07PUR9	07pur9 anopheles g
5	36	85.7	209	2 06LFL4	06lfl4 plasmodium
6	36	85.7	313	2 09P127	09p127 homo sapien
7	36	85.7	318	2 06FZ38	06fz38 bartonella
8	36	85.7	318	2 06G2J1	06g2j1 bartonella
9	36	85.7	318	2 08F290	08f290 brucella su
10	36	85.7	323	2 08LJTS	08ljts vibriophaga
11	36	85.7	332	2 08YIM3	08yim3 brucella me
12	36	85.7	360	1 MDH PYRHO	059028 pyrococcus
13	36	85.7	524	2 0934Z7	0934z7 salmonella
14	36	85.7	529	2 074YF4	074yf4 yersinia pe
15	36	85.7	529	2 06SAP7	06sap7 yersinia pe
16	36	85.7	529	2 068740	068740 yersinia pe
17	36	85.7	529	2 07ARE9	07are9 yersinia pe
18	35	83.3	129	2 09PAK0	09pak0 xylella fas
19	35	83.3	410	2 073E87	073e87 bacillus ce
20	35	83.3	496	2 08S814	08s814 bradyrhizob
21	35	83.3	10820	2 07SDK2	07sdk2 neosporea
22	35	81.0	70	2 06N0C1	06n0c1 magnetospir
23	34	81.0	77	2 086NS2	086ns2 drosophila
24	34	81.0	91	2 08RTY3	08rty3 methanopyru
25	34	81.0	107	2 06Z315	06z315 oryza sativ
26	34	81.0	140	2 08GHJ5	08ghj5 thermus the
27	34	81.0	140	2 07ZHF1	07zhf1 thermus the
28	34	81.0	147	2 08PAR9	08par9 xanthomonas
29	34	81.0	177	2 093CR0	093cr0 shigella bo
30	34	81.0	182	2 088XG7	088xg7 lactobacill
31	34	81.0	189	2 09X8Y9	09x8y9 streptomyc

32	34	81.0	193	2 06ME89	06me89 paracitamyd
33	34	81.0	207	2 06XIZ7	06xiz7 drosophila
34	34	81.0	213	2 08ZSF9	08zsf9 anabaena sp
35	34	81.0	255	2 09VQAO	09vqao drosophila
36	34	81.0	335	2 07V101	07v101 prochloroco
37	34	81.0	342	2 08TFP9	08tfp9 methanosarc
38	34	81.0	348	2 06LYA4	06lya4 methanococ
39	34	81.0	362	1 MDH PYRAB	09v05 pyrococcus
40	34	81.0	376	2 07X930	07x930 prunus aviu
41	34	81.0	382	2 06XP43	06xp43 acetivibrio
42	34	81.0	384	2 07VS20	07vs20 bordetella
43	34	81.0	384	2 07WEC2	07wec2 bordetella
44	34	81.0	439	1 IDG1 DROVA	08mx40 drosophila
45	34	81.0	439	2 Q12607	Q12607 emericella

## ALIGNMENTS

RESULT 1  
YB00 DEIRA STANDARD; PRT; 300 AA.  
AC 09RVY3;  
DT 28-FEB-2003 (Rel. 41, Created)  
DT 28-FEB-2003 (Rel. 41, Last sequence update)  
DT 25-OCT-2004 (Rel. 45, Last annotation update)  
DE Hypothetical UPF0176 protein DR1100.  
GN OrderedLocustNames=DR1100;  
OS Deinococcus radiodurans.  
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;  
OC Deinococcaceae; Deinococcus.  
OX NCBI\_Taxid=1259;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=RI / ATCC 13939 / DSM 20539 / NCBI 9279;  
RX MEDLINE=20036896; PubMed=10567266; DOI=10.1126/science.286.5444.1571;  
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,  
RA Dodson R.U., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,  
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,  
RA Vamathevan J.V., Lam P., McDonald L.A., Utterback T.R., Zalewski C.,  
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,  
RA Ketchum K.A., Nelson K.E., Salzberg S.L., Smith H.O., Venter J.C.,  
RA Fraser C.M.;  
RT "Genome sequence of the radioresistant bacterium Deinococcus  
RT radiodurans RI.";  
RL Science 286:1571-1577(1999).  
CC -!- SIMILARITY: Belongs to the UPF0176 family.  
CC -!- SIMILARITY: Contains 1 rhodanese domain.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@sib-sib.ch](mailto:license@sib-sib.ch)).  
CC  
CC EMBL; AE001960; AAF10674.1; -.  
CC PIR; G75436; G75436.  
CC TIGR; DR1100; -.  
DR HAMAP; MF\_00469; -; 1.  
DR InterPro; IPR001792; Acylphosphatase.  
DR InterPro; IPR001763; Rhodanese-like.  
DR Pfam; PF00581; Rhodanese; 1.  
DR SMART; SM00450; RHOD; 1.  
DR PROSITE; PS50206; RHODANES3; 1.  
KW Complete proteome; Hypothetical protein.  
FT DOMAIN 128 222 Rhodanese.  
SQ SEQUENCE 300 AA; 33238 MW; B33D2B8C0DC43804 CRC64;

Query Match 90.5%; Score 38; DB 1; Length 300;  
Best Local Similarity 75.0%; Pred. No. 32;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8  
 DB 261 RAHPOFEE 268

## RESULT 2

08CUD2 PRELIMINARY; PRT; 704 AA.  
 AC 08CUD2, 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Guanylyl cyclase alpha 1 subunit.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RA Nakamura I., Suzuki N.,  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB096020; BAC24016.1; -.  
 DR HSSP; P30803; IASZ.  
 DR GO; GO:0004383; F:guanylate cyclase activity; IEA.  
 DR GO; GO:0016829; F:lyase activity; IEA.  
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.  
 DR InterPro; IPR001054; G\_cyclase.  
 DR SMART; PF00211; Guanylate\_cyc; 1.  
 DR Pfam; PF00044; CYCC; 1.  
 DR PROSITE; PS00452; GUANYLATE\_CYCLASES\_1; 1.  
 DR PROSITE; PS50125; GUANYLATE\_CYCLASES\_2; 1.  
 KW lyase.  
 SQ SEQUENCE 704 AA; 79188 MW; F73FC97B68544E6A CRC64;

Query Match 88.1%; Score 37; DB 2; Length 704;  
 Best Local Similarity 85.7%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPDPEK 8  
 DB 698 SHPOFEK 704

## RESULT 3

08CH90 PRELIMINARY; PRT; 743 AA.  
 AC 08CH90, 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE Soluble guanylyl cyclase alpha 2 subunit E219G mutant.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC TISSUE=Kidney;  
 RA Nakamura I., Yao Y., Suzuki N.,  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB097860; BAC44887.1; -.  
 DR HSSP; P30803; IASZ.  
 DR GO; GO:0004383; F:guanylate cyclase activity; IEA.  
 DR GO; GO:0016829; F:lyase activity; IEA.  
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.  
 DR InterPro; IPR001054; G\_cyclase.  
 DR Pfam; PF00211; Guanylate\_cyc; 1.  
 DR SMART; SM00044; CYCC; 1.  
 DR PROSITE; PS00452; GUANYLATE\_CYCLASES\_1; 1.  
 DR PROSITE; PS50125; GUANYLATE\_CYCLASES\_2; 1.  
 KW lyase.

SQ SEQUENCE 743 AA; 83251 MW; ACF5C53E0982813A CRC64;  
 Query Match 88.1%; Score 37; DB 2; Length 743;  
 Best Local Similarity 85.7%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPDPEK 8  
 DB 737 SHPOFEK 743

## RESULT 4

07PUR9 PRELIMINARY; PRT; 2058 AA.  
 AC 07PUR9, 01-MAR-2004 (TREMBlrel. 26, Created)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 DE ENSANGP0000008445 (Fragment).  
 GN Name=ENSANGP0000008445;  
 OS Anopheles gambiae str. PE8T.  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
 OX NCBI\_TaxID=180454;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=PE8T;  
 RA Anopheles Genome Sequencing Consortium;  
 RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.  
 CC -! CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AAA801008987; EAA01115.2; -.  
 DR HSSP; Q24573; IC20.  
 DR GO; GO:0005622; C:intracellular; IEA.  
 DR GO; GO:0003677; F:DNA binding; IEA.  
 DR InterPro; IPR02086; Aldhyd\_denydrog.  
 DR InterPro; IPR001606; ARID.  
 DR Pfam; PF0388; ARID; 1.  
 DR PROSITE; PS00687; ALDEHYDE\_DEHYDR\_GLU; UNKNOWN\_1.  
 FT NON TER 1  
 FT TER 2058  
 SQ SEQUENCE 2058 AA; 214230 MW; BC18A6EFD0763F66 CRC64;

Query Match 88.1%; Score 37; DB 2; Length 2058;  
 Best Local Similarity 75.0%; Pred. No. 4.5e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8  
 DB 922 RHHPDPEK 929

## RESULT 5

06LFL4 PRELIMINARY; PRT; 209 AA.  
 AC 06LFL4, 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 DE Hypothetical protein.  
 GN ORFNames=MA16P1.55; PPF0255C;  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.

FX PubMed=12368867; DOI=10.1038/nature01095;  
 RA Hall N., Pain A., Berriman M., Churcher C., Harris B., Harris D.,  
 RA Mungall K., Bowman S., Actin R., Baker S., Barton A., Brooks K.,  
 RA Buckee C.O., Burrows C., Cherevach I., Chillingworth C.,  
 RA Chillingworth T., Christodoulou Z., Clark L., Clark R., Corto C.,  
 RA Cronin A., Davies R., Davies P., Dear P., Dearden F., Doggett J.,

RA Feltwell T., Goble A., Goodhead I., Gwilliam R., Hamlin N., Hance Z.,  
 RA Harper D., Hauser H., Hornsby T., Holroyd S., Horrocks P.,  
 RA Humphrey S., Jagels K., James D., Johnson D., Kernhorn A., Knight A.,  
 RA Konfortov B., Keyes S., Larke N., Lawson D., Lennard N., Line A.,  
 RA Maddison M., McLean J., Mooney P., Moule S., Murphy L., Oliver K.,  
 RA Ormond D., Price C., Quail M.A., Rabinowitsch E., Rajandream M.A.,  
 RA Rutter S., Rutherford K.M., Sanders M., Simmonds M., Seeger K.,  
 RA Sharp S., Smith R., Squares R., Squares S., Stevens K., Taylor K.,  
 RA Tivey A., Unwin L., Whitehead S., Woodward J., Sulton J.E., Craig A.,  
 RA Newbold C., Barrett B.G.  
 RT "Sequence of Plasmodium falciparum chromosomes 1, 3-9 and 13."  
 RL Nature 419:527-531(2002).  
 RN (12)  
 RP SEQUENCE FROM N.A.  
 RA Cherevach I., Davis P., Goodhead I., Stevens K., Mungall K.,  
 RA Bertman M., Pain A., Hall N., Atkin R., Chillingworth C., Doggett J.,  
 RA Ormond D., Sanders M., Hayes R., Hall S., Quail M., Barrett B.G.  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: CR383398; CAG25222.1; -  
 DR GO: GO:0003998; F:acetylphosphatase activity; IEA.  
 DR InterPro: IPR001792; Acylphosphatase.  
 DR Pfam: PF00708; Acylphosphatase; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 209 AA; 25373 MW; F1682AB63776B3A2 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 209;  
 Best Local Similarity 62.5%; Pred. No. 53;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RXHPOFEK 8  
 Db 44 KGHPEFEK 51

RESULT 6  
 QP127 PRELIMINARY; PRT; 313 AA.  
 AC QP127;  
 DT 01-OCT-2000 (TREMBlrel. 15, Created)  
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 DE HOM-TES-85 tumor antigen (DJ237H22.1).  
 GN Name=dx237H22.1;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
 OC NCBI\_TaxId=9606;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA TISSUE=Testis;  
 RX MEDLINE=22028805; PubMed=12032826; DOI=10.1038/sj.onc.1205491;  
 RA Tureci O., Sahin U., Koslowski M., Bues B., Bell C., Ballweber P.,  
 RA Zwick C., Eberle T., Zuber M., Villena-Heinsen C., Seltz G.,  
 RA Pfeundschn M.;  
 RT "A novel tumour associated leucine zipper protein targeting to sites  
 RT of gene transcription and splicing."  
 RL Oncogene 21:3879-3888(2002).  
 RN (12)  
 RP SEQUENCE FROM N.A.  
 RA Wilson S.;  
 RL Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: AF124430; AAF28870.1; -  
 DR EMBL: AL109751; CAC09922.1; -  
 SQ SEQUENCE 313 AA; 35937 MW; E041911D9BA1DC8B CRC64;

Query Match 85.7%; Score 36; DB 2; Length 313;  
 Best Local Similarity 62.5%; Pred. No. 84;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
 QY 1 RXHPOFEK 8  
 Db 182 RHHPOFEK 189

RESULT 7  
 Q6F238 PRELIMINARY; PRT; 318 AA.  
 AC Q6F238;  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 DE Hypothetical protein.  
 GN OrderedLocustNames=B009510;  
 OS Bartonella quintana (Rochalimaea quintana).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Bartonellaceae; Bartonella.  
 OC NCBI\_TaxId=803;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=Toulouse;  
 RX PubMed=15210978; DOI=10.1073/pnas.0305659101;  
 RA Alsmark U.C.M., Frank A.C., Karlberg E.O., Legault B.-A., Ardeli D.H.,  
 RA Canbaeck B., Eriksson A.-S., Naeslund A.K., Handley S.A., Huvet M.,  
 RA La Scala B., Holmberg M., Andersson S.G.E.;  
 RT "The louse-borne human pathogen Bartonella quintana is a genomic  
 RT derivative of the zoonotic agent Bartonella henselae."  
 RL Proc. Natl. Acad. Sci. U.S.A. 101:9716-9721(2004).  
 DR EMBL: BX897700; CAF26428.1; -  
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO: GO:0006118; F:electron transport; IEA.  
 DR InterPro: IPR000683; GFO/IDH/MOCA\_N.  
 DR InterPro: IPR004104; GFO/IDH/MOCA\_C.  
 DR Pfam: PF01408; GFO\_IDH\_MOCA; 1.  
 DR Pfam: PF02894; GFO\_IDH\_MOCA\_C; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 318 AA; 35962 MW; 638383119CD75700 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 318;  
 Best Local Similarity 75.0%; Pred. No. 85;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 RXHPOFEK 8  
 Db 121 RHHPOFEK 128

RESULT 8  
 Q6G2J1 PRELIMINARY; PRT; 318 AA.  
 AC Q6G2J1;  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 DE Hypothetical protein.  
 GN OrderedLocustNames=BH12100;  
 OS Bartonella henselae (Rochalimaea henselae).  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Bartonellaceae; Bartonella.  
 OC NCBI\_TaxId=38323;  
 RN (1)  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=ATCC 49882 / Houston 1;  
 RX PubMed=15210978; DOI=10.1073/pnas.0305659101;  
 RA Alsmark U.C.M., Frank A.C., Karlberg E.O., Legault B.-A., Ardeli D.H.,  
 RA Canbaeck B., Eriksson A.-S., Naeslund A.K., Handley S.A., Huvet M.,  
 RA La Scala B., Holmberg M., Andersson S.G.E.;  
 RT "The louse-borne human pathogen Bartonella quintana is a genomic  
 RT derivative of the zoonotic agent Bartonella henselae."  
 RL Proc. Natl. Acad. Sci. U.S.A. 101:9716-9721(2004).  
 DR EMBL: BX897699; CAF26428.1; -  
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO: GO:0006118; F:electron transport; IEA.  
 DR InterPro: IPR000683; GFO/IDH/MOCA\_N.  
 DR InterPro: IPR004104; GFO\_IDH\_MOCA\_C.  
 DR Pfam: PF01408; GFO\_IDH\_MOCA; 1.  
 DR Pfam: PF02894; GFO\_IDH\_MOCA\_C; 1.

KM Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 318 AA; 35582 MW; 7AE81771FF655355 CRC64;  
 Query Match 85.7%; Score 36; DB 2; Length 318;  
 Best Local Similarity 75.0%; Pred. No. 85;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8  
 |||||  
 121 RFHPAFEK 128

Db 121 RFHPAFEK 128

RESULT 9  
 Q8FZ90 PRELIMINARY; PRT; 318 AA.  
 ID Q8FZ90  
 AC Q8FZ90  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE Oxidoreductase, Gfo/Idh/Moca family.  
 GN OrderedLocustNames=BR1602;  
 OS Brucella suis.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Brucellaceae; Brucella.  
 OX NCBI\_TaxID=29461;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=1330 / Biovar 1;  
 RX MEDLINE=22277711; PubMed=12271122; DOI=10.1073/pnas.192319099;  
 RA Paulsen I.T., Seshadri R., Nelson K.E., Bisen J.A., Heidelberg J.F.,  
 RA Read T.D., Dodson R.U., Umayam L.A., Brinkac L.M., Beanan M.J.,  
 RA Daugherty S.C., Debey R.T., Durkin A.S., Kolonay J.F., Madupu R.,  
 RA Nelson W.C., Ayodeji B., Krahl M., Shetty J., Malek J.A.,  
 RA Van Aken S.E., Riedmiller S., Tettein H., Gill S.R., White O.,  
 RA Salzberg S.L., Hoover D.L., Lindler L.E., Hailing S.M., Boyle S.M.,  
 RA Frazer C.M.;  
 RT "The Brucella suis genome reveals fundamental similarities between  
 RT animal and plant pathogens and symbionts."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:13148-13153(2002).  
 DR EMBL: AE014453; AN30507.1; -.  
 DR TTGR: BR1602; -.  
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO: GO:0006118; P:electron transport; IEA.  
 DR InterPro: IPR000683; GFO/IDH/MOCA\_N.  
 DR InterPro: IPR004104; GFO\_IDH\_MOCA\_C.  
 DR Pfam: PF01408; GFO\_IDH\_MOCA\_C; 1.  
 DR Pfam: PF02894; GFO\_IDH\_MOCA\_C; 1.  
 KM Complete proteome.  
 SQ SEQUENCE 318 AA; 35700 MW; 5367908586945BD5 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 318;  
 Best Local Similarity 75.0%; Pred. No. 85;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8  
 |||||  
 121 RFHPAFEK 128

Db 121 RFHPAFEK 128

RESULT 10  
 Q8LTT5 PRELIMINARY; PRT; 323 AA.  
 ID Q8LTT5  
 AC Q8LTT5  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE 5' to 3' exonuclease-like protein.  
 GN Vibriophage VPV262.  
 OS Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;  
 OC T7-like viruses.  
 OX NCBI\_TaxID=194802;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RX MEDLINE=22667917; PubMed=12781722;  
 RA Hardies S.C., Comeau A.M., Serwer P., Suttle C.A.;  
 RT "The complete sequence of marine bacteriophage VPV262 infecting vibrio  
 RT parahaemolyticus indicates that an ancestral component of a T7 viral  
 RT supergroup is widespread in the marine environment."  
 RL Virology 310:359-371(2003).  
 DR EMBL: AY095314; AM28378.1; -.  
 DR GO: GO:0003677; F:DNA binding; IEA.  
 DR GO: GO:0004527; F:exonuclease activity; IEA.  
 DR InterPro: IPR008918; 5\_3\_exo\_C.  
 DR InterPro: IPR000513; Exo\_N\_I.  
 KM Exonuclease.  
 SQ SEQUENCE 323 AA; 37725 MW; 81CB297839B33951 CRC64;

Query Match 85.7%; Score 36; DB 2; Length 323;  
 Best Local Similarity 62.5%; Pred. No. 87;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8  
 |||||  
 76 KKHPEFEK 83

Db 76 KKHPEFEK 83

RESULT 11  
 Q8YIM3 PRELIMINARY; PRT; 332 AA.  
 ID Q8YIM3  
 AC Q8YIM3  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
 DE OXIDOREDUCTASE (EC 1.1.1.-).  
 GN OrderedLocustNames=BME10420;  
 OS Brucella melitensis.  
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Brucellaceae; Brucella.  
 OX NCBI\_TaxID=29459;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=16W / ATCC 23456 / Biotype 1;  
 RX MEDLINE=20020109; PubMed=11756688; DOI=10.1073/pnas.221575398;  
 RA Ivanova N., Anderson I., Bhattacharya A., Lykidis A., Reznik G.,  
 RA Jablonki L., Larsen N., D'Souza M., Bernal A., Mazur M., Goltsman E.,  
 RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Letesson J.-C.,  
 RA Haefliger R., Kyriades N.C., Overbeek R.;  
 RT "The genome sequence of the facultative intracellular pathogen  
 RT Brucella melitensis."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).  
 DR EMBL: AE009485; AAL51601.1; -.  
 DR PIR: AF3304; AF3304.  
 DR GO: GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO: GO:0006118; P:electron transport; IEA.  
 DR Pfam: PF01408; GFO\_IDH\_MOCA\_C; 1.  
 DR Pfam: PF02894; GFO\_IDH\_MOCA\_C; 1.  
 KM Complete proteome; Oxidoreductase.  
 SQ SEQUENCE 332 AA; 37530 MW; CA57412174171CDB CRC64;

Query Match 85.7%; Score 36; DB 2; Length 332;  
 Best Local Similarity 75.0%; Pred. No. 89;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8  
 |||||  
 135 RFHPAFEK 142

Db 135 RFHPAFEK 142

RESULT 12  
 MDH\_PYRHO STANDARD; PRT; 360 AA.  
 ID MDH\_PYRHO  
 AC 059028;  
 DT 30-MAY-2000 (Rel. 39, Created)  
 DT 30-MAY-2000 (Rel. 39, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)

```

DE Malate dehydrogenase (EC 1.1.1.37).
GN Name=mdh; OrderedLocNames=PH2177;
OS Pyrococcus horikoshii;
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OX NCBI_TaxID=53953;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OT3.
RX MEDLINE=9834437; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kotsugi H., Haseyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Shizuya H., Nakamura Y., Robb F.T., Horikoshi K.,
RA Mauch Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
CC -1- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.
CC -1- SUBUNIT: Homodimer (by similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the LDH2/MDH2 oxidoreductase family.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; AF000005; BAA30380.1; -.
DR PIR; B71073; B71073.
DR InterPro; IPR003767; Idb 2.
DR Pfam; PF02615; Idb 2; 1.
DR Complete proteome; NAD: Oxidoreductase; Tricarboxylic acid cycle.
SQ SEQUENCE 360 AA; 39751 MW; 6E9D8B16ECDBE66F CRC64;

Query Match
Best Local Similarity 85.7%; Score 36; DB 1; Length 360;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8
DB 311 RKHPFER 318

RESULT 13
Q93427 PRELIMINARY; PRT; 524 AA.
ID Q93427;
AC Q93427;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Hypothetical protein HCM2.0069c.
GN OrderedLocNames=HCM2.0069c;
OS Salmonella typhi.
OG Plasmid pHCM2.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Salmonella.
OX NCBI_TaxID=601;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CT18;
RX MEDLINE=21534947; PubMed=11677608; DOI=10.1038/35101607;
RA Parthill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
RA Churcher C.W., Mungall K.L., Bentley S.D., Holden N.T.G., Sebatina M.,
RA Baker S., Basham D., Brooks K., Chillingworth T., Conerton P.,
RA Cronin A., Davis P., Davies R.M., Dow L., White N., Farrar J.,
RA Felwell T., Hamlin N., Haque A., Hien T.T., Holtroyd S., Jagels K.,
RA Krogh A., Larsen T.S., Leather S., O'Gaora P., Parry C.,
RA Quail M.A., Rutherford K.M., Simmonds M., Skelton J., Stevens K.,

```

```

RA Whitehead S., Barrell B.G.;
RT "Complete genome sequence of a multiple drug resistant Salmonella
RT enterica serovar Typhi CT18.";
RL Nature 413:848-852(2001).
DR EMBL; AL513384; CAD09936.1; -.
KW Complete proteome; Hypothetical protein; Plasmid.
SQ SEQUENCE 524 AA; 60431 MW; 99EC4CDD42F793F7 CRC64;

Query Match
Best Local Similarity 85.7%; Score 36; DB 2; Length 524;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8
DB 14 RKHPFER 21

RESULT 14
Q74YP4 PRELIMINARY; PRT; 529 AA.
ID Q74YP4;
AC Q74YP4;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Hypothetical protein.
GN OrderedLocNames=PMT024;
OS Yersinia pestis.
OG Plasmid pMT1.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=91001 / Biovar Mediaevalis;
RA Song Y., Tong Z., Wang L., Han Y., Zhang J., Pei D., Wang J., Zhou D.,
RA Han Y., Pang X., Zhai J., Chen F., Qin H., Wang J., Li S., Guo Z.,
RA Ye C., Du Z., Lin W., Wang J., Yu J., Yang H., Wang J., Huang P.,
RA Yang R.;
RL Submitted (APR-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; AE017045; AAS0659.1; -.
KW Hypothetical protein; Plasmid.
SQ SEQUENCE 529 AA; 61074 MW; 9AF9D1141768A0BB CRC64;

Query Match
Best Local Similarity 85.7%; Score 36; DB 2; Length 529;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFEK 8
DB 19 RKHPFER 26

RESULT 15
Q65AP7 PRELIMINARY; PRT; 529 AA.
ID Q65AP7;
AC Q65AP7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein.
GN Yersinia pestis.
OG Plasmid pG8786.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Yersinia.
OX NCBI_TaxID=632;
RN [1]
RP SEQUENCE FROM N.A.
RC PubMed=15385458;
RA Golubov A., Neubauer H., Nolting C., Heesemann J., Rakin A.;
RA "Structural Organization of the P1ra Virulence-Associated Plasmid of
RT Rhinoceros-Positive Yersinia pestis.";
RL Infect. Immun. 72:5613-5621(2004).
DR EMBL; AJ698720; CAG27426.1; -.

```

KW Hypothetical protein; Plasmid.  
SQ SEQUENCE 529 AA; 61074 MW; 9AE89400496940BB CRC64;  
Query Match 85.7%; Score 36; DB 2; Length 529;  
Best Local Similarity 62.5%; Pred. No. 1.5e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
QY 1 RXHPQFEK 8  
|:|:|:  
|:|:|:  
Db 19 RRHPEYEK 26

Search completed: March 2, 2005, 12:44:19  
Job time : 30.7236 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 36.0325 Seconds  
(without alignments)  
85.869 Million cell updates/sec

Title: SEQ8  
Perfect score: 42  
Sequence: 1 rxhpqfek 8

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	88.1	8	AAW59212	AAW59212 Streptavi
2	37	88.1	8	AA835433	AA835433 Epitope p
3	37	88.1	8	AA835442	AA835442 Nascent p
4	37	88.1	8	AA868616	AA868616 Strep-Tag
5	37	88.1	8	AAO19065	AAO19065 Mutation
6	37	88.1	8	AB857464	AB857464 STREP tag
7	37	88.1	8	AB877486	AB877486 AAV-helpe
8	37	88.1	8	AB873584	AB873584 Streptagi
9	37	88.1	8	AAE37230	AAE37230 Strept-tag
10	37	88.1	8	ABP60361	ABP60361 Streptavi
11	37	88.1	8	ABP60368	ABP60368 Streptavi
12	37	88.1	8	ADA09808	ADA09808 Streptag e
13	37	88.1	8	AD884588	AD884588 Streptavi
14	37	88.1	8	AD885500	AD885500 Streptavi
15	37	88.1	8	AD829930	AD829930 Antibody-
16	37	88.1	8	AD006951	AD006951 StrepTag
17	37	88.1	8	ADP90832	ADP90832 Protein/p
18	37	88.1	8	AD820244	AD820244 Strep tag
19	37	88.1	8	ADR72509	ADR72509 Dehydroge
20	37	88.1	9	AB831054	AB831054 Peptide p
21	37	88.1	10	AA197662	AA197662 Influenza
22	37	88.1	10	AA899027	AA899027 Streptavi
23	37	88.1	10	AAU80475	AAU80475 Peptide S
24	37	88.1	10	ABP56623	ABP56623 C-termina
25	37	88.1	10	ADP11064	ADP11064 Strep-tag

26	37	88.1	10	ADN16967	ADN16967 Human res
27	37	88.1	11	AD026489	AD026489 Strep-tag
28	37	88.1	11	AA838373	AA838373 Epitope t
29	37	88.1	11	AD820228	AD820228 Strep tag
30	37	88.1	19	ABG74881	ABG74881 Bacteriop
31	37	88.1	19	ABG74882	ABG74882 Bacteriop
32	37	88.1	19	ABG74880	ABG74880 Bacteriop
33	37	88.1	19	AD139157	AD139157 Construct
34	37	88.1	19	AD139155	AD139155 Construct
35	37	88.1	19	AD139156	AD139156 Construct
36	37	88.1	24	ABP60362	ABP60362 Streptavi
37	37	88.1	35	ABP60369	ABP60369 Streptavi
38	37	88.1	36	ABP60370	ABP60370 Streptavi
39	37	88.1	36	ABP60370	ABP60370 Streptavi
40	37	88.1	117	AAU97558	AAU97558 Synthetic
41	37	88.1	117	AAU97553	AAU97553 Synthetic
42	37	88.1	117	AAU97557	AAU97557 Synthetic
43	37	88.1	117	AAU97559	AAU97559 Synthetic
44	37	88.1	117	AAU97555	AAU97555 Synthetic
45	37	88.1	117	AAU97556	AAU97556 Synthetic

## ALIGNMENTS

RESULT 1  
ID AAW59212  
AAW59212 standard; peptide; 8 AA.

AC AAW59212;

DT 27-AUG-1998 (first entry)

DE Streptavidin tagged peptide ligand #2.

KW Streptavidin; ligand; binding affinity; mutant; isolation; purification;  
recovery; immobilise.

OS Synthetic.

PN EP835934-A2.

PD 15-APR-1998.

PF 09-OCT-1997; 97EP-00117504.

PR 10-OCT-1996; 96DE-01041876.

PA (BIOA-) INST BIOANALYTIK GMBH.

PI Skerra A, Voss S;

DR WPI; 1998-218868/20.

PT Streptavidin mutants with higher binding affinity for peptide ligands -  
have mutation in amino acid region 44-53, used to isolate, purify or  
determine fusion proteins including these ligands.

PS Claim 10; Page 11; 21pp; German.

XX AAW59211 and AAW59212 are ligands used in a method to assay binding  
affinity of streptavidin mutants. These mutants have a mutation within  
the amino acid (aa) region 44-53 of the wild-type protein show a higher  
binding affinity than the wild-type for peptide ligands that include the  
sequence of formula Trp-X-His-Pro-Gln-Phe-Y-Z where X = any aa; Y and Z  
are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant streptavidin  
mutants can be used to isolate, purify and determine proteins or to  
determine/recover substances that contain streptavidin-binding groups.  
CC Such compounds may also be used to immobilise fusions on microtitre  
plates, microbeads or sensor chips  
XX Sequence 8 AA;

Query Match 88.1%; Score 37; DB 2; Length 8;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFEK 8  
 :|||||  
 DB 2 SHPQFEK 8

## RESULT 2

AAB35433  
 ID AAB35433 standard; peptide; 8 AA.

XX AAB35433;  
 AC AAB35433;  
 XX

DT 23-MAY-2001 (first entry)  
 DE Eptlope peptide #3.  
 XX

XX Nascent protein detection; protein analysis; aminoacylated tRNA;  
 KW BODIPY marker; disease diagnosis.  
 XX

XX Unidentified.  
 OS WO200114578-A1.  
 PN

XX WO200114578-A1.  
 PD

XX 01-MAR-2001.  
 PF

XX 23-AUG-2000; 2000WO-US023233.  
 PR

XX 25-AUG-1999; 99US-00382736.  
 PR

XX 25-AUG-1999; 99US-00382950.  
 XX

XX (AMBE-) AMBERGEN INC.  
 PA

XX Rothschild KJ, Gite S, Olejnik J;  
 PI

XX WPI; 2001-168972/17.  
 DR

XX Method for detecting nascent proteins by fluorescence comprises  
 PT misaminoacylating a tRNA molecule with a marker compound, useful for  
 PT detecting mutations in proteins, e.g. cancer.  
 PS

XX Disclosure; Page 47; 204pp; English.  
 XX

XX The present invention describes a method of detecting nascent proteins  
 CC involving aminoacylating a tRNA molecule with a 4',4-difluoro-4-bora-3a,4A  
 CC -diaz-a-s-indacene (BODIPY) marker leading to the production of a  
 CC misaminoacylated tRNA. This enables the detection, isolation and analysis  
 CC of nascent proteins using UV without the usual accompanying radioactivity  
 CC problems. It may be used to detect mutations, for example in cancer.  
 CC Duchenne muscular dystrophy, adenomatous polyposis coli and colon cancer  
 XX

XX Sequence 8 AA;  
 SQ

Query Match 88.1%; Score 37; DB 4; Length 8;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFEK 8  
 :|||||  
 DB 2 SHPQFEK 8

XX AAB35442  
 ID AAB35442 standard; peptide; 8 AA.  
 XX

XX AAB35442;  
 AC AAB35442;  
 XX

XX 23-MAY-2001 (first entry)  
 DT

XX Nascent protein detection method related peptide #4.  
 DE

XX Nascent protein detection; protein analysis; aminoacylated tRNA;  
 KW BODIPY marker; disease diagnosis.  
 KM

XX Unidentified.  
 OS WO200114578-A1.  
 PN

XX WO200114578-A1.  
 PD

XX 01-MAR-2001.  
 PF

XX 23-AUG-2000; 2000WO-US023233.  
 PR

XX 25-AUG-1999; 99US-00382736.  
 PR

XX 25-AUG-1999; 99US-00382950.  
 XX

XX (AMBE-) AMBERGEN INC.  
 PA

XX Rothschild KJ, Gite S, Olejnik J;  
 PI

XX WPI; 2001-168972/17.  
 DR

XX Method for detecting nascent proteins by fluorescence comprises  
 PT misaminoacylating a tRNA molecule with a marker compound, useful for  
 PT detecting mutations in proteins, e.g. cancer.  
 PS

XX Example 22; Page 153; 204pp; English.  
 XX

XX The present invention describes a method of detecting nascent proteins  
 CC involving aminoacylating a tRNA molecule with a 4',4-difluoro-4-bora-3a,4A  
 CC -diaz-a-s-indacene (BODIPY) marker leading to the production of a  
 CC misaminoacylated tRNA. This enables the detection, isolation and analysis  
 CC of nascent proteins using UV without the usual accompanying radioactivity  
 CC problems. It may be used to detect mutations, for example in cancer,  
 CC Duchenne muscular dystrophy, adenomatous polyposis coli and colon cancer  
 XX

XX Sequence 8 AA;  
 SQ

Query Match 88.1%; Score 37; DB 4; Length 8;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFEK 8  
 :|||||  
 DB 2 SHPQFEK 8

XX AAB68616  
 ID AAB68616 standard; peptide; 8 AA.  
 XX

XX AAB68616;  
 AC AAB68616;  
 XX

XX 27-APR-2001 (first entry)  
 DT

XX Strep-Tag II sequence.  
 DE

XX Protein-RNA fusion; Strep-Tag II.  
 KW

XX Unidentified.  
 OS WO200107657-A1.  
 PN

XX WO200107657-A1.  
 PD

XX 01-FEB-2001.  
 PF

XX 19-JUL-2000; 2000WO-US019653.  
 PR

XX 27-JUL-1999; 99US-0145834P.  
 PR

XX (PHYL-) PHYLLOS INC.  
 PA

XX Kurz M, Lohse P, Wagner R;  
 PI

XX WPI; 2001-182803/18.  
 DR

XX Affixing a peptide acceptor to an RNA molecule useful for producing  
PT fusion proteins for isolating proteins or nucleic acids with desired  
PT properties through attachment of a peptide acceptor to the 3' end of an  
PT RNA molecule.  
XX  
PS Example 5; Page 22; 56pp; English.  
XX  
CC The present invention relates to a method for affixing a peptide acceptor  
CC to an RNA molecule through the formation of a covalent bond, noncovalent  
CC bond, or by chemical ligation. The method is useful for producing RNA-  
CC protein fusions which can be used for the isolation of proteins or  
CC nucleic acids with desired properties from large pools of partially or  
CC completely random amino acid or nucleic acid sequences. The present  
CC sequence is a Strep-tag II used in the present invention  
XX  
SQ Sequence 8 AA;  
XX  
Query Match 88.1%; Score 37; DB 4; Length 8;  
Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 2 XHPDPEK 8  
:|||||  
Db 2 SHPDPEK 8  
XX  
RESULT 5  
AA019065  
ID AA019065 standard; peptide; 8 AA.  
XX  
AC AA019065;  
XX  
DT 14-NOV-2002 (first entry)  
XX  
DE Mutation detection method tag peptide SEQ ID NO: 34.  
XX  
KM Mutation detection; primer; mutant; tag; tumour suppressor gene;  
KM protein production; cancer.  
XX  
OS Synthetic.  
XX  
PN WO20026675-A2.  
XX  
PD 29-AUG-2002.  
XX  
PF 15-FEB-2002; 2002WO-EP001651.  
XX  
PR 16-FEB-2001; 2001DE-01007317.  
XX  
PA (PLAC ) MAX PLANCK GES FOERDERUNG WISSENSCHAFTEN.  
XX  
PI Kahmann S, Mueller O;  
XX  
DR MPI; 2002-674959/72.  
XX  
N-PSDB; AAL49464.  
XX  
PT Detecting mutations in nucleic acid, useful for diagnosis and  
PT characterization of tumors, by amplification, in vitro transcription and  
PT translation, then protein detection.  
XX  
PS Disclosure; Fig 5; 62pp; German.  
XX  
CC The present invention relates to a method of detecting mutations in a  
CC nucleic acid by amplifying the nucleic acid to produce a double-stranded  
CC amplicon, in vitro transcription and translation of this amplicon, and  
CC detection of the translated protein. The primers used for amplification  
CC are designed to produce an amplicon that is translatable and allows  
CC differentiation between translation products of wild-type and mutated  
CC nucleic acids. The method is used to detect mutations in tumour  
CC suppressor genes, for (early) diagnosis, monitoring and characterisation  
CC of tumours (especially of bladder and intestines) and in the germ line  
CC (using nucleic acids from embryos or blood cells). A new multi-tag vector

CC is used to detect or verify the reading frame of a nucleic acid cloned in  
CC it, and to determine the suitability of detectable peptides for analysis  
CC and/or purification of a recombinant protein, expressed from a sequence  
CC cloned in the vector. The present sequence is a tag peptide which was  
CC used in the invention.  
XX  
SQ Sequence 8 AA;  
XX  
Query Match 88.1%; Score 37; DB 5; Length 8;  
Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 2 XHPDPEK 8  
:|||||  
Db 2 SHPDPEK 8  
XX  
RESULT 6  
ABB57464  
ID ABB57464 standard; peptide; 8 AA.  
XX  
AC ABB57464;  
XX  
DT 18-MAR-2002 (first entry)  
XX  
DE STREP tag II peptide.  
XX  
KM Immunomodulatory human MHC class II antigen-binding protein; HLA;  
KM human leukocyte antigen; immune system; immunosuppression; antibody;  
KM major histocompatibility complex; antirheumatic; antiarthritic;  
KM neuroprotective; antiinflammatory; antidiabetic; antipsoriatic;  
KM immunosuppressive; dermatological; antichyroid; nephrotropic; psoriasis;  
KM chryomimetic; hepatotropic; immune response suppressor; narcolepsy;  
KM rheumatoid arthritis; juvenile arthritis; multiple sclerosis; Insulitis;  
KM Grave's disease; insulin-dependent diabetes; Hashimoto's disease;  
KM systemic lupus erythematosus; ankylosing spondylitis; myasthenia gravis;  
KM transplant rejection; graft versus host disease; pemphigus vulgaris;  
KM glomerulonephritis; thyroiditis; pancreatitis; primary biliary cirrhosis;  
KM irritable bowel disease; Sjogren's syndrome.  
XX  
OS Synthetic.  
XX  
PN WO200187338-A1.  
XX  
PD 22-NOV-2001.  
XX  
PF 14-MAY-2001; 2001WO-US015626.  
XX  
PR 12-MAY-2000; 2000EP-00110063.  
XX  
PR 06-OCT-2000; 2000US-0238762P.  
XX  
PA (GPCB-) GPC BIOTECH AG.  
XX  
PA (MORP-) MORPHOSYS AG.  
XX  
PI Nagy Z, Tesar M, Thomassen-Wolf E;  
XX  
DR MPI; 2002-075289/10.  
XX  
PT Composition for suppressing immune response, treating diseases of immune  
PT system, has polypeptide comprising antibody-based antigen-binding domain  
PT of human composition, which binds antigen expressed on a cell surface.  
XX  
PS Disclosure; Page 31; 139pp; English.  
XX  
CC The present invention describes a composition (I), comprising a  
CC polypeptide comprising an antibody-based antigen-binding domain of human  
CC composition with binding specificity for an antigen expressed on the  
CC surface of a cell, where treating cells expressing the antigen with the  
CC polypeptides leads to suppression of an immune response, and the IC50 for  
CC the suppression of immune response is 1 microm or less. (I) has  
CC antirheumatic, antiarthritic, neuroprotective, antiinflammatory,  
CC antidiabetic, antipsoriatic, immunosuppressive, dermatological,  
CC antichyroid, nephrotropic, chryomimetic and hepatotropic activities, and

CC can be used as a suppressor of immune response. (1) is useful for  
 CC suppressing activation or proliferation of a cell of the immune system,  
 CC suppressing IL-2 secretion by a cell, the interaction of a cell of the  
 CC immune system with another cell, immunosuppressing a patient and for  
 CC killing a cell expressing an antigen, human leukocyte antigen (HLA)-DR on  
 CC the surface of the cell, where neither cytotoxic entities nor  
 CC immunological mechanisms are needed to cause or lead to the killing. (1)  
 CC (optionally linked to cytotoxic or immunogenic agent) is useful for  
 CC preparing a pharmaceutical preparation for the treatment of rheumatoid  
 CC arthritis, juvenile arthritis, multiple sclerosis, Grave's disease,  
 CC insulin-dependent diabetes, narcolepsy, psoriasis, systemic lupus  
 CC erythematosus, ankylosing spondylitis, transplant rejection, graft versus  
 CC host disease, Hashimoto's disease, myasthenia gravis, pemphigus vulgaris,  
 CC glomerulonephritis, thyroiditis, pancreatitis, insulinitis, primary biliary  
 CC cirrhosis, irritable bowel disease and Sjogren's syndrome in humans.  
 CC ABA92469 to ABA92474 and ABB57457 to ABB57590 represent sequence used in  
 CC the exemplification of the present invention

CC Sequence 8 AA;

Query Match 88.1%; Score 37; DB 5; Length 8;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8  
 :|||||  
 DB 2 SHPQFEK 8

RESULT 7  
 ABB77486  
 ID ABB77486 standard; peptide; 8 AA.

XX ABB77486;  
 XX 22-JUL-2002 (first entry)

DE AAV-helper plasmid related Strep-peptide tag SEQ ID NO 17.

KW Adeno-associated virus; AAV; glycoprotein; ITR; cap; rep; vector.

OS Synthetic.  
 PN WO200238782-A2.

PD 16-MAY-2002.

PF 13-NOV-2001; 2001WO-EP013125.

PR 13-NOV-2000; 2000DE-01056210.

PA (ARIM-) ARIMEDES BIOTECHNOLOGY GMBH.

PI Orberger G, Hellmuth K, Wagener C;

DR MPI; 2002-435853/46.

PT Vector system for preparing recombinant adeno-associated viral particles,  
 PT used for high-level expression of heterologous therapeutic proteins in  
 PT eukaryotic cells.

PS Disclosure; Page 17; 59pp; German.

CC The invention relates to a viral vector system for preparing recombinant  
 CC adeno-associated virus (AAV) particles comprising; at least two plasmid  
 CC vectors (AB158983 and AB158984) that include the two inverted terminal  
 CC repeats (ITR) of AAV and additional sequences and plasmid vectors without  
 CC ITRs but containing the rep and cap genes of AAV required for replication  
 CC and packaging. The system is useful for producing recombinant AAV for  
 CC production of a wide range of therapeutic glycoproteins in eukaryotic  
 CC cells. The system provides efficient, large scale production of  
 CC heterologous proteins in mammalian cells, without requiring an adenovirus  
 CC helper. It is not toxic to host cells and does not cause lysis, so

CC produced proteins are highly pure. The present sequence is that of a  
 CC peptide tag, useful to the invention

CC Sequence 8 AA;

Query Match 88.1%; Score 37; DB 5; Length 8;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8  
 :|||||  
 DB 2 SHPQFEK 8

RESULT 8  
 ABG73584  
 ID ABG73584 standard; protein; 8 AA.

XX ABG73584;  
 XX 03-MAR-2003 (first entry)

DE StreptagII affinity peptide.

KW Non-protein L-amino acid; O-acetyl-L-serine; agrochemical production;  
 KW nucleophile; O-acetyl-L-serine sulphydrylase; pharmaceutical production.

OS Synthetic.  
 PN EP1247869-A1.

PD 09-OCT-2002.

PF 28-MAR-2002; 2002EP-00007262.

PR 04-APR-2001; 2001DE-01016881.

PR 03-MAY-2001; 2001DE-01021515.

PA (CONE ) CONSORTIUM ELEKTROCHEM IND GMBH.

PI Maier T, Gaebert C;

DR WPI; 2003-077522/08.  
 PT Production of non-protein L-amino acids useful for the manufacture of  
 PT pharmaceuticals and agrochemicals, comprises an enzyme catalyzed reaction  
 PT of O-acetyl-L-serine with a nucleophile.

PS Example 1; Page 8; 20pp; German.

CC This invention describes a novel method for the production of non-protein  
 CC L-amino acids by the reaction of O-acetyl-L-serine with a nucleophile in  
 CC the presence of a catalyst comprising O-acetyl-L-serine sulphydrylase at  
 CC pH 5-7.4. The method of the invention is useful for the manufacture of  
 CC pharmaceuticals and agrochemicals. In contrast to the process described  
 CC in DE 10046934, a high nucleophile concentration can be used which  
 CC includes toxic compounds. This sequence represents an affinity peptide  
 CC containing a StreptagII motif which is used in the construction of fusion  
 CC genes containing Bacitracin colli cy8K and cy8M fragments

CC Sequence 8 AA;

Query Match 88.1%; Score 37; DB 6; Length 8;  
 Best Local Similarity 85.7%; Pred. No. 1.8e+06;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8  
 :|||||  
 DB 2 SHPQFEK 8

RESULT 9  
 AAE37230

```

ID AAE37230 standard; peptide; 8 AA.
XX
AC AAE37230;
XX
DT 07-AUG-2003 (first entry)
XX
DE Strep-tag II epitope peptide.
XX
KW Gene expression; therapy; isolation; epitope.
XX
OS Synthetic.
XX
PN WO2003038049-A2.
XX
PD 08-MAY-2003.
XX
PF 29-OCT-2002; 2002WO-US034645.
XX
PR 29-OCT-2001; 2001US-0340689P.
XX
PA (RENO-) RENOVIS INC.
PI Helmut N, Serafini TA, Shyjan AW;
DR WPI; 2003-430512/40.
XX
PT Isolating cell-type specific mRNAs, useful in gene expression analysis or
PT quantification in a specific cell in a heterogeneous cell mixture, by
PT isolating mRNA bound to ribosomes or proteins that bind mRNA in a cell
PT type specific manner.
XX
PS Example 2; Page 121; 136pp; English.
XX
CC The invention relates to a method for isolating mRNA from a population of
CC cells. The method involves selectively isolating ribosomes or proteins
CC that bind mRNA in a cell type specific manner and then isolating the mRNA
CC bound to the ribosomes or proteins that bind mRNA. The method is useful
CC for facilitating the analysis and quantification of gene expression in a
CC selected cell type present within a heterogeneous cell mixture. The
CC method may also be used in diagnostics or therapies for human diseases.
CC The present sequence is Strep-tag II epitope peptide. This sequence is
CC used to illustrate the method of the invention
XX
SQ Sequence 8 AA:
Query Match 88.1%; Score 37; DB 6; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 XHPQEX 8
DB 2 SHPQEX 8
RESULT 10
ID ABE60361 standard; peptide; 8 AA.
XX
AC ABE60361;
XX
DT 28-MAR-2003 (first entry)
XX
DE Streptavidin 11 tag peptide SEQ ID NO 2.
XX
KW Streptavidin; protein chip; microtitre plate; detection.
XX
OS Synthetic.
XX
PN DE10113776-A1.
XX
PD 02-OCT-2002.
XX
PF 21-MAR-2001; 2001DE-01013776.
XX

```

```

XX
PR 21-MAR-2001; 2001DE-01013776.
XX
PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
PI Schmidt T;
XX
DR WPI; 2003-031166/03.
XX
PT New isolated peptide, useful as affinity purification tag for recombinant
PT protein, comprises at least two high-affinity streptavidin-binding
PT modules.
XX
PS Example 1; Page 8; 18pp; German.
XX
CC The invention relates to an isolated peptide (I) comprising at least two
CC individual modules separated by 0-50 amino acids, with each containing at
CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
CC streptavidin binding modules, are useful as affinity handles for
CC purification of recombinant fusion proteins (FP), also for detecting FP,
CC e.g. on protein chips or microtitre plates. The modules in (I) bind
CC strongly to streptavidin, with a co-operative effect that provides
CC stronger binding than a single tag, but are displaced by a competitor.
CC (I) does not interfere with the function of attached proteins (II) (so it
CC may not be essential to remove it); facilitates detection and has easily
CC controllable binding properties. (I) is particularly used for purifying
CC FP from dilute solution in batch formats (which use simpler apparatus
CC than column methods and result in lower loss of FP). The present sequence
CC is that of a streptavidin tag peptide disclosed with the invention
XX
SQ Sequence 8 AA:
Query Match 88.1%; Score 37; DB 6; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.8e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
QY 2 XHPQEX 8
DB 2 SHPQEX 8
RESULT 11
ID ABE60368 standard; peptide; 8 AA.
XX
AC ABE60368;
XX
DT 28-MAR-2003 (first entry)
XX
DE Streptavidin binding peptide SEQ ID NO 9.
XX
KW Streptavidin; protein chip; microtitre plate; detection.
XX
OS Synthetic.
XX
PN DE10113776-A1.
XX
PD 02-OCT-2002.
XX
PF 21-MAR-2001; 2001DE-01013776.
XX
PR 21-MAR-2001; 2001DE-01013776.
XX
PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
PI Schmidt T;
XX
DR WPI; 2003-031166/03.
XX
PT New isolated peptide, useful as affinity purification tag for recombinant
PT protein, comprises at least two high-affinity streptavidin-binding
PT modules.
XX

```

PS Claim 5; Page 16; 18pp; German.  
XX  
CC The invention relates to an isolated peptide (I) comprising at least two  
CC individual modules separated by 0-50 amino acids, with each containing at  
CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain  
CC streptavidin binding modules, are useful as affinity handles for  
CC purification of recombinant fusion proteins (FP), also for detecting FP,  
CC e.g. on protein chips or microtitre plates. The modules in (I) bind  
CC strongly to streptavidin, with a co-operative effect that provides  
CC stronger binding than a single tag, but are displaced by a competitor.  
CC (I) does not interfere with the function of attached proteins (II) (so it  
CC may not be essential to remove it); facilitates detection and has easily  
CC controllable binding properties. (I) is particularly used for purifying  
CC FP from dilute solution in batch formats (which use simpler apparatus  
CC than column methods and result in lower loss of FP). The present sequence  
CC is that of a streptavidin binding peptide disclosed with the invention  
XX  
SQ Sequence 8 AA;  
XX  
Query Match 88.1%; Score 37; DB 6; Length 8;  
Best Local Similarity 85.7%; Pred. No. 1.8e+06; Indels 0; Gaps 0;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 2 XHPQFEK 8  
DB :|||||  
2 SHPQFEK 8  
XX  
RESULT 12  
ADA09808  
ID ADA09808 standard; peptide; 8 AA.  
XX  
AC ADA09808;  
XX  
DT 06-NOV-2003 (first entry)  
XX  
DE Streptag epitope useful as an affinity marker.  
XX  
KW Non-radioactive marker; nascent protein detection method;  
XX cellular translation system; cell-free translation system;  
XX diethylmethanoboron difluoride dye;  
XX 4,4-difluoro-4-bora-3a,4a-diaza-s-indacene dye; vaccine; drug;  
XX human disease screening; human disorder; protein separation;  
XX affinity marker; Streptag epitope.  
XX  
OS Synthetic.  
XX  
PN US2003092031-A1.  
XX  
PD 15-MAY-2003.  
XX  
PF 18-JUN-2002; 2002US-00174368.  
XX  
PR 25-AUG-1999; 99US-00382736.  
XX 23-AUG-2000; 2000WO-US023233.  
XX 21-JUN-2002; 2002US-00049332.  
XX  
PA (AMBE-) AMBERGEN INC.  
XX  
PI Rothschild KJ, Gite S, Olejnik J;  
XX  
DR WPI; 2003-576764/54.  
XX  
PT Detecting, analyzing or isolating nascent proteins comprises introducing  
XX a modified nucleic acid template into a cellular or cell-free translation  
XX system to generate a nascent protein having at least an N-terminal  
XX marker.  
XX  
PS Disclosure; Page 15; 76pp; English.  
XX  
CC The present invention relates to non-radioactive markers used in the  
XX detection and analysis of nascent proteins translated in cellular or cell  
XX -free translation systems. The preferred non-radioactive markers are

CC diethylmethanoboron difluoride (4,4-difluoro-4-bora-3a,4a-diaza-s-  
CC indacene) dyes. The detection method of the invention is a gel-free  
CC method that comprises introducing a modified nucleic acid template into a  
CC translation system under conditions such that a nascent protein is  
CC generated, the protein comprising at least an N-terminal marker. The  
CC method is useful in detecting, analyzing and isolating nascent proteins  
CC produced in a cell-free or cellular translation system without the use of  
CC radioactive amino acids or other radioactive labels. Compositions  
CC comprising nascent proteins translated in the presence of markers may be  
CC used as vaccines or as drugs for humans and other animals. The method and  
CC a kit containing reagents for the detection of nascent proteins may be  
CC used as a rapid means to screen humans or other animals for the presence  
CC of certain diseases or disorders. The present sequence represents an  
CC epitope that can be used as an affinity marker for protein separation.  
XX  
SQ Sequence 8 AA;  
XX  
Query Match 88.1%; Score 37; DB 7; Length 8;  
Best Local Similarity 85.7%; Pred. No. 1.8e+06; Indels 0; Gaps 0;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
OY 2 XHPQFEK 8  
DB :|||||  
2 SHPQFEK 8  
XX  
RESULT 13  
ADB84588  
ID ADB84588 standard; peptide; 8 AA.  
XX  
AC ADB84588;  
XX  
DT 04-DEC-2003 (first entry)  
XX  
DE Streptavidin conserved peptide #2.  
XX  
KW cell-free transcription system; cell-free translation system;  
XX protein synthesis; matrix; streptavidin.  
XX  
OS Escherichia coli.  
XX  
PN DE10137792-A1.  
XX  
PD 27-FEB-2003.  
XX  
PF 06-AUG-2001; 2001DE-01037792.  
XX  
PR 06-AUG-2001; 2001DE-01037792.  
XX  
PA (ERDM/) ERDMANN V.  
XX  
PI Erdmann VA, Lamla T, Stiege W;  
XX  
DR WPI; 2003-343999/33.  
XX  
PT Expressing genes in cell-free system, useful for preparation of proteins,  
XX comprises that the protein formed is removed from solution by binding to  
XX a matrix.  
XX  
PS Claim 13; Col 8; 8pp; German.  
XX  
CC This invention describes a novel method of expressing genes in a cell-  
CC free transcription and translation system which comprises using a  
CC reaction solution containing all necessary components of the  
CC transcription/translation system, amino acids, nucleotides and  
CC metabolites that supply energy and that are needed for synthesis. The  
CC proteins formed are immobilised on a matrix. The method allows simple  
CC recovery of proteins without a separate isolation step and the amount of  
CC proteins produced can be determined before a reaction is complete.  
CC Continuous removal of proteins prevents it interfering with the  
CC expression process, making possible synthesis of proteins that interact  
CC adversely with the process, so normally produced only in very low yields.  
CC The use of a matrix concentrates the proteins produced and is applicable



Qy 2 XHPQPEK 8  
:|||||  
Db 2 SHPQPEK 8

Search completed: March 2, 2005, 13:02:46  
Job time : 37.0325 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

## OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 24.878 Seconds  
(without alignments)  
105.489 Million cell updates/sec

Title: SEQ8

Perfect score: 42

Sequence: 1 rxhpfek 8

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubppaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubppaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubppaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubppaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubppaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubppaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubppaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubppaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubppaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubppaa/US09C\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubppaa/US10B\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubppaa/US10B\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubppaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubppaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubppaa/US10D\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/2/pubppaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/2/pubppaa/US11\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/2/pubppaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	38	90.5	36	14	US-10-026-578B-11
2	37	88.1	8	9	US-09-809-517A-9
3	37	88.1	8	9	US-09-809-517A-9
4	37	88.1	8	9	US-09-809-517A-9
5	37	88.1	8	12	US-09-813-145-7
6	37	88.1	8	13	US-10-104-218-5
7	37	88.1	8	13	US-10-104-218-5
8	37	88.1	8	13	US-10-208-357-9
9	37	88.1	8	14	US-10-001-934-8
10	37	88.1	8	14	US-10-026-578B-2
11	37	88.1	8	14	US-10-026-578B-9
12	37	88.1	8	14	US-10-174-368A-7
13	37	88.1	8	14	US-10-264-127-8
14	37	88.1	8	15	US-10-339-712-8
15	37	88.1	8	15	US-10-339-712-67

14	37	88.1	8	15	US-10-275-046-4	Sequence 4, Appli
15	37	88.1	8	15	US-10-425-000-76	Sequence 76, Appli
16	37	88.1	8	15	US-10-424-999-25	Sequence 25, Appli
17	37	88.1	8	15	US-10-358-283-23	Sequence 23, Appli
18	37	88.1	8	16	US-10-628-432-41	Sequence 41, Appli
19	37	88.1	8	17	US-10-494-248-17	Sequence 17, Appli
20	37	88.1	8	17	US-10-634-645-11	Sequence 11, Appli
21	37	88.1	8	17	US-10-719-523-8	Sequence 8, Appli
22	37	88.1	9	9	US-09-983-067-3	Sequence 3, Appli
23	37	88.1	10	9	US-09-809-517A-6	Sequence 6, Appli
24	37	88.1	10	15	US-10-147-211A-20	Sequence 20, Appli
25	37	88.1	11	15	US-10-354-983-29	Sequence 29, Appli
26	37	88.1	11	16	US-10-628-432-25	Sequence 25, Appli
27	37	88.1	21	9	US-09-809-517A-30	Sequence 30, Appli
28	37	88.1	22	9	US-09-809-517A-33	Sequence 33, Appli
29	37	88.1	24	9	US-09-809-517A-31	Sequence 31, Appli
30	37	88.1	24	14	US-10-026-578B-3	Sequence 3, Appli
31	37	88.1	25	9	US-10-026-578B-4	Sequence 4, Appli
32	37	88.1	25	9	US-09-809-517A-34	Sequence 34, Appli
33	37	88.1	36	14	US-10-026-578B-10	Sequence 10, Appli
34	37	88.1	117	10	US-09-977-137A-4	Sequence 4, Appli
35	37	88.1	117	10	US-09-977-137A-5	Sequence 5, Appli
36	37	88.1	117	10	US-09-977-137A-7	Sequence 7, Appli
37	37	88.1	117	10	US-09-977-137A-8	Sequence 8, Appli
38	37	88.1	117	10	US-09-977-137A-9	Sequence 9, Appli
39	37	88.1	117	10	US-09-977-137A-10	Sequence 10, Appli
40	37	88.1	117	10	US-09-977-137A-11	Sequence 11, Appli
41	37	88.1	117	10	US-09-977-137A-12	Sequence 12, Appli
42	37	88.1	118	10	US-09-977-137A-6	Sequence 6, Appli
43	37	88.1	245	17	US-10-887-228A-1	Sequence 1, Appli
44	37	88.1	246	17	US-10-887-228A-9	Sequence 9, Appli
45	37	88.1	252	17	US-10-887-228A-5	Sequence 5, Appli

## ALIGNMENTS

RESULT 1  
US-10-026-578B-11  
; Sequence 11, Application US/10026578B  
; Publication No. US20030083474A1  
; GENERAL INFORMATION:  
; APPLICANT: IBA (GmbH)  
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta  
; FILE REFERENCE: 100810.01US1  
; CURRENT APPLICATION NUMBER: US/10/026,578B  
; CURRENT FILING DATE: 2002-11-11  
; PRIOR APPLICATION NUMBER: DE 101 13 776.1  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: PCT/EP01/11846  
; PRIOR FILING DATE: 2001-10-12  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 11  
; LENGTH: 36  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic Peptide  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (9)..(28)  
; OTHER INFORMATION: Of the five (GGGS) repeats between position 9 and 28, one or more  
; OTHER INFORMATION: repeats may be missing. However, at least one (GGGS) repeat will  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Artificial Sequence represents peptide binding module  
US-10-026-578B-11  
Query Match 90.5%; Score 38; DB 14; Length 36;  
Best Local Similarity 75.0%; Pred. No. 11;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPQFEK 8

DB 29 QSHPOFEK 36

## RESULT 2

US-09-809-517A-9  
; Sequence 9, Application US/09809517A  
; Patent No. US20020034733A1  
; GENERAL INFORMATION:  
; APPLICANT: Lohning, Corinna  
; TITLE OF INVENTION: No. US20020034733A1e1 methods for displaying (poly)peptides/prote  
; TITLE OF INVENTION: particles via disulfide bonds  
; FILE REFERENCE: MORPHO/11  
; CURRENT APPLICATION NUMBER: US/09/809,517A  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: EP 99114072.4  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: EP 00103551.8  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 9  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module  
US-09-809-517A-9

Query Match 88.1%; Score 37; DB 9; Length 8;

Best Local Similarity 85.7%; Pred. No. 1.2e+06; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8

DB 2 SHPOFEK 8

## RESULT 3

US-09-973-145-7  
; Sequence 7, Application US/09973145  
; Patent No. US20020132240A1  
; GENERAL INFORMATION:  
; APPLICANT: Rothschild, Kenneth J.  
; APPLICANT: Gile, Sadanand  
; APPLICANT: Olejnik, Jerzy  
; TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins  
; FILE REFERENCE: AMBER-06819  
; CURRENT APPLICATION NUMBER: US/09/973,145  
; PRIOR FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: 09/382,950  
; PRIOR FILING DATE: 1999-08-25  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Synthetic  
US-09-973-145-7

Query Match 88.1%; Score 37; DB 9; Length 8;

Best Local Similarity 85.7%; Pred. No. 1.2e+06; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8

DB 2 SHPOFEK 8

## RESULT 4

US-09-813-197-8  
; Sequence 8, Application US/09813197  
; Publication No. US2005009013A1  
; GENERAL INFORMATION:  
; APPLICANT: Rothschild, Kenneth  
; APPLICANT: Gile, Sadanand  
; APPLICANT: Olejnik, Jerzy  
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Prot  
; FILE REFERENCE: AMBER-03951  
; CURRENT APPLICATION NUMBER: US/09/813,197  
; CURRENT FILING DATE: 2001-03-20  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 8  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
US-09-813-197-8

Query Match 88.1%; Score 37; DB 12; Length 8;

Best Local Similarity 85.7%; Pred. No. 1.2e+06; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8

DB 2 SHPOFEK 8

## RESULT 5

US-10-104-218-5  
; Sequence 5, Application US/10104218  
; Publication No. US20020177196A1  
; GENERAL INFORMATION:  
; APPLICANT: MAIER, Thomas  
; APPLICANT: GABBERT, Carsten  
; TITLE OF INVENTION: PROCESS FOR PREPARING NON-PROTEINOGENIC L-AMINO ACIDS  
; FILE REFERENCE: MAIER, T. ET AL.-2  
; CURRENT APPLICATION NUMBER: US/10/104,218  
; PRIOR FILING DATE: 2002-03-22  
; PRIOR APPLICATION NUMBER: US/02/03-22  
; PRIOR FILING DATE: 2001-04-04  
; PRIOR APPLICATION NUMBER: German No. US20020177196A1 101 16 881.0  
; PRIOR FILING DATE: 2001-05-03  
; NUMBER OF SEQ ID NOS: 5  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 5  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; NAME/KEY: PEPTIDE  
; LOCATION: (1)-(8)  
; OTHER INFORMATION: StreptagII affinity peptide for protein purification  
US-10-104-218-5

Query Match 88.1%; Score 37; DB 13; Length 8;

Best Local Similarity 85.7%; Pred. No. 1.2e+06; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPQFEK 8

DB 2 SHPOFEK 8

## RESULT 6

US-10-208-357-9

```
; Sequence 9, Application US/10208357
; Publication No. US20020182687A1
; GENERAL INFORMATION:
; APPLICANT: Kurz, Markus
; APPLICANT: Lohse, Peter
; APPLICANT: Wagner, Richard
; TITLE OF INVENTION: Peptide Acceptor Ligation Methods
; FILE REFERENCE: 50036/031002
; CURRENT APPLICATION NUMBER: US/10/208,357
; CURRENT FILING DATE: 2002-07-30
; PRIOR APPLICATION NUMBER: US/09/619,103
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/145,834
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: designed sequence to act as an identifying tag
US-10-208-357-9

Query Match      88.1%; Score 37; DB 13; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQEX 8
       :|||||
Db      2 SHPQEX 8

RESULT 7
US-10-001-934-8
; Sequence 8, Application US/10001934
; Publication No. US20030032782A1
; GENERAL INFORMATION:
; APPLICANT: NADY, ZOLTAN
; TITLE OF INVENTION: HUMAN PEPTIDES/PROTEINS CAUSING OR LEADING TO THE
; TITLE OF INVENTION: KILLING OF CELLS INCLUDING LYMPHOID TUMOR CELLS
; FILE REFERENCE: GPCG-P01-003
; CURRENT APPLICATION NUMBER: US/10/001,934
; CURRENT FILING DATE: 2001-11-15
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-001-934-8

Query Match      88.1%; Score 37; DB 14; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQEX 8
       :|||||
Db      2 SHPQEX 8

RESULT 8
US-10-026-578B-2
; Sequence 2, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
```

```
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 2
; LENGTH: 8
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; NAME/KEY: misc feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-2

Query Match      88.1%; Score 37; DB 14; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQEX 8
       :|||||
Db      2 SHPQEX 8

RESULT 9
US-10-026-578B-9
; Sequence 9, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; NAME/KEY: misc feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-9

Query Match      88.1%; Score 37; DB 14; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQEX 8
       :|||||
Db      2 SHPQEX 8

RESULT 10
US-10-174-368A-7
; Sequence 7, Application US/10174368A
; Publication No. US20030092031A1
; GENERAL INFORMATION:
; APPLICANT: Rothchild, Kenneth J.
; APPLICANT: Gile, Sadanand
; APPLICANT: Olejnik, Jerzy
```

```

; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Prot
; FILE REFERENCE: AMBER-07145
; CURRENT APPLICATION NUMBER: US/10/174,368A
; CURRENT FILING DATE: 2002-06-18
; PRIOR APPLICATION NUMBER: 10/049,332
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: PCT/US00/23233
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/382,736
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
; NAME/KEY: misc_feature
; OTHER INFORMATION: Synthetic
US-10-174-368A-7

Query Match      88.1%; Score 37; DB 14; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFEK 8
       :|||||
Db      2 SHPQFEK 8

RESULT 11
US-10-264-127-8
; Sequence 8, Application US/10264127
; Publication No. US20030190643A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth
; APPLICANT: Olejnik, Jerzy
; APPLICANT: Gile, Sadanand
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Prot
; FILE REFERENCE: AMBER-03951
; CURRENT APPLICATION NUMBER: US/10/264,127
; CURRENT FILING DATE: 2002-10-03
; PRIOR APPLICATION NUMBER: US/09/382,736B
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-264-127-8

Query Match      88.1%; Score 37; DB 14; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFEK 8
       :|||||
Db      2 SHPQFEK 8

RESULT 12
US-10-339-712-8
; Sequence 8, Application US/10339712
; Publication No. US20040014071A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Gile, Sadanand
; APPLICANT: Olejnik, Jerzy

```

```

; APPLICANT: Lim, Mark
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent
; FILE REFERENCE: AMBER-07199
; CURRENT APPLICATION NUMBER: US/10/339,712
; CURRENT FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: 10/049,332
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: PCT/US00/23233
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/382,736
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-339-712-8

Query Match      88.1%; Score 37; DB 15; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFEK 8
       :|||||
Db      2 SHPQFEK 8

RESULT 13
US-10-339-712-67
; Sequence 67, Application US/10339712
; Publication No. US20040014071A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth J.
; APPLICANT: Olejnik, Jerzy
; APPLICANT: Lim, Mark
; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent
; FILE REFERENCE: AMBER-07199
; CURRENT APPLICATION NUMBER: US/10/339,712
; CURRENT FILING DATE: 2003-01-10
; PRIOR APPLICATION NUMBER: 10/049,332
; PRIOR FILING DATE: 2002-06-21
; PRIOR APPLICATION NUMBER: PCT/US00/23233
; PRIOR FILING DATE: 2000-08-23
; PRIOR APPLICATION NUMBER: 09/382,736
; PRIOR FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 67
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-10-339-712-67

Query Match      88.1%; Score 37; DB 15; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFEK 8
       :|||||
Db      2 SHPQFEK 8

RESULT 14
US-10-275-046-4
; Sequence 4, Application US/10275046

```

```

; Publication No. US20040019187A1
; GENERAL INFORMATION:
; APPLICANT: Nagy et al.
; TITLE OF INVENTION: IMMUNOMODULATORY HUMAN MHC CLASS II ANTIGEN-BINDING POLYPEPTIDES
; FILE REFERENCE: GPCG-P01-260
; CURRENT APPLICATION NUMBER: US/10/275,046
; CURRENT FILING DATE: 2002-10-31
; PRIOR APPLICATION NUMBER: 00110063.5
; PRIOR FILING DATE: 2000-05-12
; NUMBER OF SEQ ID NOS: 97
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 4
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: STREP tag II
US-10-275-046-4

```

```

Query Match      88.1%; Score 37; DB 15; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy      2 XHPQFEK 8
        :|||||
        2 SHPQFEK 8

```

```

RESULT 15
US-10-425-000-76
; Sequence 76, Application US/10425000
; Publication No: US20040052777A1

```

```

; GENERAL INFORMATION:
; APPLICANT: Nebbit, Mark
; APPLICANT: C&Aeron, Beatrice
; APPLICANT: Blanche, Francis
; TITLE OF INVENTION: K-1 single Polypeptides and Methods for Using Them to Inhibit
; TITLE OF INVENTION: Angiogenesis
; FILE REFERENCE: ST01027-B
; CURRENT APPLICATION NUMBER: US/10/425,000
; CURRENT FILING DATE: 2003-04-29
; PRIOR APPLICATION NUMBER: 10/233,675
; PRIOR FILING DATE: 2002-09-04
; NUMBER OF SEQ ID NOS: 105
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO: 76
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Purification tag
US-10-425-000-76

```

```

Query Match      88.1%; Score 37; DB 15; Length 8;
Best Local Similarity 85.7%; Pred. No. 1.2e+06;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Oy      2 XHPQFEK 8
        :|||||
        2 SHPQFEK 8

```

```

Search completed: March 2, 2005, 14:18:51
Job time : 24.878 secs

```

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 9.10569 Seconds  
(without alignments)  
65.585 Million cell updates/sec

Title: SEQ8  
Perfect score: 42  
Sequence: 1 rxhpqfek 8

Scoring table: BLOSUM62X  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A COMB pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B COMB pep:\*  
3: /cgn2\_6/prodata/1/1aa/5A COMB pep:\*  
4: /cgn2\_6/prodata/1/1aa/5B COMB pep:\*  
5: /cgn2\_6/prodata/1/1aa/5A COMB pep:\*  
6: /cgn2\_6/prodata/1/1aa/5B COMB pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	37	88.1	8	US-08-948-097-2	Sequence 2, Appl
2	37	88.1	8	US-09-382-950-7	Sequence 7, Appl
3	37	88.1	8	US-09-382-7358-8	Sequence 8, Appl
4	37	88.1	8	US-09-619-103-9	Sequence 9, Appl
5	37	88.1	8	US-10-104-218-5	Sequence 9, Appl
6	37	88.1	8	US-09-809-517A-9	Sequence 9, Appl
7	37	88.1	10	US-09-809-517A-6	Sequence 9, Appl
8	37	88.1	21	US-09-809-517A-30	Sequence 30, Appl
9	37	88.1	22	US-09-809-517A-33	Sequence 33, Appl
10	37	88.1	24	US-09-809-517A-31	Sequence 31, Appl
11	37	88.1	25	US-09-809-517A-34	Sequence 34, Appl
12	37	88.1	117	US-09-977-137A-5	Sequence 5, Appl
13	37	88.1	117	US-09-977-137A-7	Sequence 5, Appl
14	37	88.1	117	US-09-977-137A-7	Sequence 8, Appl
15	37	88.1	117	US-09-977-137A-8	Sequence 9, Appl
16	37	88.1	117	US-09-977-137A-9	Sequence 10, Appl
17	37	88.1	117	US-09-977-137A-11	Sequence 11, Appl
18	37	88.1	117	US-09-977-137A-12	Sequence 12, Appl
19	37	88.1	118	US-09-977-137A-6	Sequence 6, Appl
20	37	88.1	118	US-09-977-137A-6	Sequence 7, Appl
21	34	81.0	45	US-08-630-915A-71	Sequence 71, Appl
22	34	81.0	45	US-09-879-957-71	Sequence 71, Appl
23	34	81.0	180	US-09-134-000C-4700	Sequence 4700, Ap
24	34	81.0	207	US-09-248-796A-22051	Sequence 22051, A
25	34	81.0	213	US-09-107-532A-3797	Sequence 3797, Ap
26	34	81.0	234	US-09-270-767-32857	Sequence 32857, A
27	34	81.0	234	US-09-270-767-48074	Sequence 48074, A

28	34	81.0	275	4	US-09-270-767-58917	Sequence 58917, A
29	34	81.0	608	4	US-09-270-767-32937	Sequence 32937, A
30	34	81.0	608	4	US-09-270-767-48154	Sequence 48154, A
31	34	81.0	712	4	US-09-248-796A-18407	Sequence 18407, A
32	34	81.0	801	4	US-09-270-767-43549	Sequence 43549, A
33	34	81.0	979	1	US-08-308-881-6	Sequence 6, Appl
34	34	81.0	979	2	US-09-058-263-6	Sequence 6, Appl
35	34	81.0	979	2	US-09-058-263-6	Sequence 6, Appl
36	34	81.0	979	3	US-09-058-264-6	Sequence 6, Appl
37	34	81.0	979	4	US-09-455-962-6	Sequence 6, Appl
38	34	81.0	979	5	PCT-US95-06530-6	Sequence 6, Appl
39	34	81.0	3666	2	US-08-222-617A-12	Sequence 12, Appl
40	34	81.0	3727	2	US-08-222-617A-27	Sequence 27, Appl
41	34	81.0	3728	2	US-08-222-617A-2	Sequence 2, Appl
42	33	78.6	114	4	US-09-270-767-42719	Sequence 42719, A
43	33	78.6	169	4	US-09-252-991A-32254	Sequence 32254, A
44	33	78.6	413	4	US-09-723-546-11	Sequence 11, Appl
45	33	78.6	482	4	US-09-107-532A-4868	Sequence 4868, Ap

## ALIGNMENTS

```
RESULT 1
US-08-948-097-2
; Sequence 2, Application US/08948097C
; Patent No. 6103493
; GENERAL INFORMATION:
; APPLICANT: Skeira, Arne
; APPLICANT: Voss, Selma
; TITLE OF INVENTION: Streptavidin Mutains
; FILE REFERENCE: HUBR 1119
; CURRENT APPLICATION NUMBER: US/08/948,097C
; EARLIER FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: DE 196 41 876.3
; EARLIER FILING DATE: 1996-10-10
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 2
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: BINDING
; OTHER INFORMATION: Binding ligand for streptavidin
US-08-948-097-2

Query Match      88.1%; Score 37; DB 3; Length 8;
Best Local Similarity 85.7%; Pred. No. 4,1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY      2 XHPQFEK 8
      :|||||
Db      2 SHPQFEK 8

RESULT 2
US-09-382-950-7
; Sequence 7, Application US/09382950
; Patent No. 6103337
; GENERAL INFORMATION:
; APPLICANT: Rothchild, Kenneth
; APPLICANT: Gite, Sedarand
; APPLICANT: Olejnik, Uerzy
; TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins
; FILE REFERENCE: AMBER-03879
; CURRENT APPLICATION NUMBER: US/09/382,950
; CURRENT FILING DATE: 1999-08-25
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 7
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial/Unknown
```

```
FEATURE:
NAME/KEY: misc feature
LOCATION: ( )..T
OTHER INFORMATION: Synthetic
US-09-382-950-7
```

```
Query Match      88.1%; Score 37; DB 3; Length 8;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQEX 8
        :|||||
Db      2 SHPQEX 8
```

```
RESULT 3
US-09-382-736B-8
Sequence 8, Application US/09382736B
Patent No. 630628
GENERAL INFORMATION:
APPLICANT: Rothschild, Kenneth
APPLICANT: Gile, Sadanand
APPLICANT: Olejnik, Jerry
TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Pro
FILE REFERENCE: AMBER-03951
CURRENT APPLICATION NUMBER: US/09/382,736B
CURRENT FILING DATE: 1999-08-25
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn version 3.0
SEQ ID NO 8
LENGTH: 8
TYPE: PRT
FEATURE:
ORGANISM: Artificial Sequence
OTHER INFORMATION: Synthetic
US-09-382-736B-8
```

```
Query Match      88.1%; Score 37; DB 3; Length 8;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQEX 8
        :|||||
Db      2 SHPQEX 8
```

```
RESULT 4
US-09-619-103-9
Sequence 9, Application US/09619103
Patent No. 6429300
GENERAL INFORMATION:
APPLICANT: Kurtz, Markus
APPLICANT: Lohse, Peter
APPLICANT: Wagner, Richard
TITLE OF INVENTION: Peptide Acceptor Ligation Methods
FILE REFERENCE: 50036/031002
CURRENT APPLICATION NUMBER: US/09/619,103
CURRENT FILING DATE: 2000-07-19
PRIOR APPLICATION NUMBER: 60/145,834
PRIOR FILING DATE: 1999-07-27
NUMBER OF SEQ ID NOS: 26
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9
LENGTH: 8
TYPE: PRT
FEATURE:
ORGANISM: Artificial Sequence
OTHER INFORMATION: designed sequence to act as an identifying tag
US-09-619-103-9
```

```
Query Match      88.1%; Score 37; DB 4; Length 8;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQEX 8
        :|||||
Db      2 SHPQEX 8
```

```
RESULT 5
US-10-104-218-5
Sequence 5, Application US/10104218
Patent No. 6579705
GENERAL INFORMATION:
APPLICANT: MAIER, Thomas
APPLICANT: GAEBERT, Carsten
TITLE OF INVENTION: PROCESS FOR PREPARING NON-PROTEINOGENIC L-AMINO ACIDS
FILE REFERENCE: MAIER, T. ET AL.-2
CURRENT APPLICATION NUMBER: US/10/104,218
CURRENT FILING DATE: 2002-03-22
PRIOR APPLICATION NUMBER: German No. 6579705 101 16 881.0
PRIOR FILING DATE: 2001-04-04
PRIOR APPLICATION NUMBER: German No. 6579705 101 21 515.0
PRIOR FILING DATE: 2001-05-03
NUMBER OF SEQ ID NOS: 5
SOFTWARE: PatentIn version 3.0
SEQ ID NO 5
LENGTH: 8
TYPE: PRT
FEATURE:
ORGANISM: Artificial Sequence
NAME/KEY: PEPTIDE
LOCATION: (1)..(8)
OTHER INFORMATION: StreptagII affinity peptide for protein purification
US-10-104-218-5
```

```
Query Match      88.1%; Score 37; DB 4; Length 8;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQEX 8
        :|||||
Db      2 SHPQEX 8
```

```
RESULT 6
US-09-809-517A-9
Sequence 9, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
CURRENT FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-16
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.0
SEQ ID NO 9
LENGTH: 8
TYPE: PRT
FEATURE:
ORGANISM: artificial sequence
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-9
```

```
Query Match      88.1%; Score 37; DB 4; Length 8;
Best Local Similarity 85.7%; Pred. No. 4.1e+05;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQEX 8
        :|||||
```



Db 2 SHPOFEK 8

RESULT 7  
US-09-809-517A-6  
; Sequence 6, Application US/09809517A  
; Patent No. 6753136  
; GENERAL INFORMATION:  
; APPLICANT: Lohning, Corinna  
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b  
; FILE REFERENCE: MORPHO/11  
; CURRENT APPLICATION NUMBER: US/09/809,517A  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: EP 99114072.4  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: EP 00103551.8  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 6  
; LENGTH: 10  
; TYPE: PRT  
; ORGANISM: artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module  
US-09-809-517A-6

Query Match 88.1%; Score 37; DB 4; Length 10;  
Best Local Similarity 85.7%; Pred. No. 0.43;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFEK 8  
:|||||  
Db 4 SHPOFEK 10

RESULT 8  
US-09-809-517A-30  
; Sequence 30, Application US/09809517A  
; Patent No. 6753136  
; GENERAL INFORMATION:  
; APPLICANT: Lohning, Corinna  
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b  
; FILE REFERENCE: MORPHO/11  
; CURRENT APPLICATION NUMBER: US/09/809,517A  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: EP 99114072.4  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: EP 00103551.8  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 30  
; LENGTH: 21  
; TYPE: PRT  
; ORGANISM: artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module  
US-09-809-517A-30

Query Match 88.1%; Score 37; DB 4; Length 21;  
Best Local Similarity 85.7%; Pred. No. 0.97;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFEK 8  
:|||||  
Db 15 SHPOFEK 21

RESULT 9  
US-09-809-517A-33

; Sequence 33, Application US/09809517A  
; Patent No. 6753136  
; GENERAL INFORMATION:  
; APPLICANT: Lohning, Corinna  
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b  
; FILE REFERENCE: MORPHO/11  
; CURRENT APPLICATION NUMBER: US/09/809,517A  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: EP 99114072.4  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: EP 00103551.8  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 33  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module  
US-09-809-517A-33

Query Match 88.1%; Score 37; DB 4; Length 22;  
Best Local Similarity 85.7%; Pred. No. 1;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFEK 8  
:|||||  
Db 16 SHPOFEK 22

RESULT 10  
US-09-809-517A-31  
; Sequence 31, Application US/09809517A  
; Patent No. 6753136  
; GENERAL INFORMATION:  
; APPLICANT: Lohning, Corinna  
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b  
; FILE REFERENCE: MORPHO/11  
; CURRENT APPLICATION NUMBER: US/09/809,517A  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: EP 99114072.4  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: EP 00103551.8  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: Patentin version 3.0  
; SEQ ID NO 31  
; LENGTH: 24  
; TYPE: PRT  
; ORGANISM: artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module  
US-09-809-517A-31

Query Match 88.1%; Score 37; DB 4; Length 24;  
Best Local Similarity 85.7%; Pred. No. 1.1;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFEK 8  
:|||||  
Db 18 SHPOFEK 24

RESULT 11  
US-09-809-517A-34  
; Sequence 34, Application US/09809517A  
; Patent No. 6753136  
; GENERAL INFORMATION:  
; APPLICANT: Lohning, Corinna  
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

```
; TITLE OF INVENTION: Particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentln Ver. 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34
```

```
Query Match      88.1%; Score 37; DB 4; Length 25;
Best Local Similarity 85.7%; Pred. No. 1.2;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQFEK 8
        :|||||
Db      19 SHPQFEK 25
```

```
RESULT 12
US-09-977-137A-4
; Sequence 4, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4
```

```
Query Match      88.1%; Score 37; DB 4; Length 117;
Best Local Similarity 85.7%; Pred. No. 6.3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQFEK 8
        :|||||
Db      111 SHPQFEK 117
```

```
RESULT 13
US-09-977-137A-5
; Sequence 5, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
```

```
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 5
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-5
```

```
Query Match      88.1%; Score 37; DB 4; Length 117;
Best Local Similarity 85.7%; Pred. No. 6.3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQFEK 8
        :|||||
Db      111 SHPQFEK 117
```

```
RESULT 14
US-09-977-137A-7
; Sequence 7, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 7
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-7
```

```
Query Match      88.1%; Score 37; DB 4; Length 117;
Best Local Similarity 85.7%; Pred. No. 6.3;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQFEK 8
        :|||||
Db      111 SHPQFEK 117
```

```
RESULT 15
US-09-977-137A-8
; Sequence 8, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentln Ver. 2.0
; SEQ ID NO 8
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
```

; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: chelon  
US-09-977-137A-8

Query Match 88.1%; Score 37; DB 4; Length 117;  
Best Local Similarity 85.7%; Pred. No. 6.3;  
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFEK 8  
:|||||  
Db 111 SHPOFEK 117

Search completed: March 2, 2005, 12:25:35  
Job time : 10.1057 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 6.11382 Seconds  
(without alignments)  
125.901 Million cell updates/sec

Title: SEQ9

Perfect score: 42

Sequence: 1 rxhpgfer 8

Scoring table: BLOSUM62DX

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 79:\*

1: p1r1:\*

2: p1r2:\*

3: p1r3:\*

4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	92.9	360	2 B71073	probable malate de
2	37	88.1	300	2 G75436	conserved hypothet
3	37	88.1	362	2 H75131	malate dehydrogena
4	35	83.3	129	2 A82547	hypothetical prote
5	35	83.3	338	2 T10538	heat shock protein
6	35	83.3	420	2 S71199	dnau protein homol
7	34	81.0	55	2 T35092	probable small hyd
8	34	81.0	142	2 B82223	hypothetical prote
9	34	81.0	189	2 T36642	hypothetical prote
10	34	81.0	213	2 AD2533	hypothetical prote
11	34	81.0	313	2 H68297	conserved hypothet
12	34	81.0	355	2 T24938	hypothetical prote
13	34	81.0	360	2 T18140	hypothetical prote
14	34	81.0	376	2 C75580	adenine deaminase-
15	34	81.0	508	2 C82138	conserved hypothet
16	34	81.0	543	2 A84134	acetolactate synth
17	34	81.0	621	2 A71516	hypothetical prote
18	34	81.0	627	2 T00484	hypothetical prote
19	34	81.0	755	2 H72262	sensor histidine k
20	34	81.0	845	2 T07052	probable potassium
21	34	81.0	918	2 A55277	hexokinase (EC 2.7
22	34	81.0	3746	1 YGPIV3	alpha-aminoadipyl-
23	34	81.0	3791	1 YGPIV8	alpha-aminoadipyl-
24	33	78.6	105	2 T00283	hypothetical prote
25	33	78.6	310	2 C84701	hypothetical prote
26	33	78.6	332	2 AF3304	oxidoreductase (EC
27	33	78.6	347	2 B85640	hypothetical prote
28	33	78.6	360	2 F83203	probable ATP-bind
29	33	78.6	452	2 F84421	hypothetical prote

30	33	78.6	465	2 B83449	conserved hypothet
31	33	78.6	467	1 KCHUN	neutrophil collage
32	33	78.6	468	2 T32586	hypothetical prote
33	33	78.6	529	2 T14947	hypothetical prote
34	33	78.6	1577	2 T15851	hypothetical prote
35	32	76.2	69	2 C82724	hypothetical prote
36	32	76.2	73	2 S40416	hexokinase C - rat
37	32	76.2	115	2 T13519	hypothetical prote
38	32	76.2	159	2 D75278	hypothetical prote
39	32	76.2	191	2 T35792	beta-glucosidase -
40	32	76.2	203	2 JQ1973	HrpA-A protein - m
41	32	76.2	215	2 AC3392	maleylpyruvate iso
42	32	76.2	250	2 T34242	hypothetical prote
43	32	76.2	256	2 T10109	tyrosin (EC 3.4.21
44	32	76.2	293	2 E82582	Dnau protein XP223
45	32	76.2	311	2 AF2668	deacetylase (impor

#### ALIGNMENTS

RESULT 1  
B71073  
probable malate dehydrogenase - Pyrococcus horikoshii  
C:Species: Pyrococcus horikoshii  
C>Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 12-Jul-2004  
C:Accession: B71073  
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin  
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Oguchi  
DNA Res. 5, 55-76, 1998  
A>Title: Complete sequence and gene organization of the genome of a hyper-thermophilic  
A:Reference number: A71000; MUID:98344137; PMID:9679194  
A:Accession: B71073  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-360 <KAM>  
A:Cross-references: UNIPROT:O59028; GB:AP000005; NID:g3236132; PIDN:BAA30380.1; PID:g325  
A:Experimental source: strain OT3  
A>Note: This accession replaces an interim accession for a sequence replaced by GenBank  
C:Genetics:  
A:Gene: PH1277

Query Match 92.9%; Score 39; DB 2; Length 360;  
Best Local Similarity 75.0%; Pred. No. 6.4;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RXHPQFER 8  
|:|:|:|  
Db 311 RKHPFER 318

RESULT 2  
G75436  
conserved hypothetical protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: G75436  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma  
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.  
Science 286, 1571-1577, 1999  
A>Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: AV5250; MUID:20036896; PMID:10567266  
A:Accession: G75436  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-300 <NH>  
A:Cross-references: UNIPROT:Q9RCV9; GB:AE001960; GB:AE000513; NID:g6458833; PIDN:AAF1067  
A:Experimental source: strain R1  
C:Genetics:  
A:Gene: DR1100  
A:Map position: 1  
C:Superfamily: Escherichia coli ycea protein

Query Match 88.1%; Score 37; DB 2; Length 300;  
 Best Local Similarity 85.7%; Pred. No. 13;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPQFER 7  
 |::|||  
 Db 261 RAHPQFER 267

## RESULT 3

H75131 malate dehydrogenase PAB1791 - *Pyrococcus abyssi* (strain Ozeay)

C:Species: *Pyrococcus abyssi*

C>Date: 20-Aug-1999 #sequence\_revision 20-Aug-1999 #text\_change 09-Jul-2004

C:Accession: H75131

R:Anonymous, Genoscope

Submitted to the EMBL Data Library, July 1999

A:Description: *Pyrococcus abyssi* genome sequence: Insights into archaeal chromosome stru

A:Reference number: A75001

A:Accession: H75131

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-362 <KAM>

A:Cross-references: UNIPROT:Q9V0D5; GB:AJ248285; GB:AL096836; NID:G5458067; PIDN:CAB4976

C:Genetics:

A:Gene: PAB1791

C:Superfamily: malate dehydrogenase ylbC

Query Match 88.1%; Score 37; DB 2; Length 362;  
 Best Local Similarity 75.0%; Pred. No. 16;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPQFER 8  
 |::|||  
 Db 311 RKHPQFER 318

## RESULT 4

A82547

hypothetical protein XF2515 (imported) - *Xylella fastidiosa* (strain 9a5c)

C:Species: *Xylella fastidiosa*

C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004

C:Accession: A82547

R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen

Nature 406, 151-157, 2000

A>Title: The genome sequence of the plant pathogen *Xylella fastidiosa*.

A:Reference number: A82515; MUID:20365717; PMID:10910347

A>Note: for a complete list of authors see reference number A59328 below

A:Accession: A82547

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-129 <SIM>

A:Cross-references: UNIPROT:Q9PAK0; GB:AE004059; GB:AE003849; NID:G9107718; PIDN:AAFE531

A:Experimental source: strain 9a5c

R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvares, R.; A

Briones, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carre, H

de-Neto, E.; Docena, C.; El-Dorri, H.; Facinanti, A.F.; Ferreira, A.J.S.

submitted to GenBank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm

J.D.; Junqueira, M.L.; Kemper, E.L.; Kitching, J.P.; Krieger, J.B.; Kuramae, E.B.; Laig

chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Martino, C.L.; Marques, M.V.; Martins, E

A:Authors: Martins, E.M.F.; Matsumura, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.

F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A

Rodrigues, V.; Rosa, A.U.; de Rosa Jr., V.B.; de Sa, R.G.; Santelli, R.V.; Sawasak

A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silveir

M.; Tsubako, M.H.; Valada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF2515

Query Match 83.3%; Score 35; DB 2; Length 129;  
 Best Local Similarity 62.5%; Pred. No. 13;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPQFER 8  
 |::|||  
 Db 94 RHHPQFER 101

## RESULT 5

T30538

heat shock protein homolog dnaJ - *Trypanosoma cruzi*

C:Species: *Trypanosoma cruzi*

C>Date: 22-Oct-1999 #sequence\_revision 22-Oct-1999 #text\_change 09-Jul-2004

C:Accession: T30538

R:Bringaud, F.; Vedrenne, C.; Cuvillier, A.; Parzy, D.; Baltz, D.; Tetaud, E.; Pays, E.;

Mol. Biochem. Parasitol. 94, 249-264, 1998

A>Title: Conserved organization of genes in *trypanosomatids*.

A:Reference number: Z16580; MUID:98418771; PMID:9747975

A:Accession: T30538

A>Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-338 <BRI>

A:Cross-references: UNIPROT:O76230; EMBL:AF031927; NID:G3452217; PID:G3452219; PIDN:AAC3

C:Genetics:

A:Gene: dnaJ

C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

F:4-70/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 83.3%; Score 35; DB 2; Length 338;  
 Best Local Similarity 62.5%; Pred. No. 35;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPQFER 8  
 |::|||  
 Db 242 KHPQFER 249

## RESULT 6

S71199

dnaJ protein homolog atj3 - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C>Date: 28-Oct-1996 #sequence\_revision 27-Feb-1997 #text\_change 09-Jul-2004

C:Accession: S71199

R:Zhou, R.; Kroczyńska, B.; Miernik, J.A.

submitted to the EMBL Data Library, March 1995

A:Description: Atj3, an *Arabidopsis thaliana* homologue of the *Escherichia coli* DnaJ.

A:Reference number: S71199

A:Accession: S71199

A:Molecule type: mRNA

A:Residues: 1-420 <ZHO>

A:Cross-references: UNIPROT:Q42530; EMBL:U22340; NID:G1872162; PIDN:ABA49030.1; PID:G727

C:Genetics:

A:Gene: atj3

C:Superfamily: heat shock protein dnaJ; dnaJ amino-terminal homology

F:14-75/Domain: dnaJ amino-terminal homology <DNJ>

Query Match 83.3%; Score 35; DB 2; Length 420;  
 Best Local Similarity 62.5%; Pred. No. 45;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPQFER 8  
 |::|||  
 Db 260 KHPQFER 267

## RESULT 7

T35092

probable small hydrophilic protein - *Streptomyces coelicolor*

C:Species: *Streptomyces coelicolor*

C>Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 09-Jul-2004

C:Accession: T35092

R:Seeger, K.U.; Harris, D.; Bentley, S.D.; Parthill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, July 1999  
 A:Reference number: Z21567  
 A:Accession: T35092  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-55 <SEE>  
 A:Cross-references: UNIPROT:Q9S286; EMBL:AL096884; PIDN:CA84410.1; GSPDB:GN00070; SCODE  
 C:Genetics:  
 A:Gene: SCODEB:SC4G6.32

Query Match  
 Best Local Similarity 81.0%; Score 34; DB 2; Length 55;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFER 8  
 :|||  
 Db 10 RXHPOAER 17

RESULT 8  
 B82223  
 hypothetical protein VC1253 [imported] - Vibrio cholerae (strain N16961 serogroup O1)  
 C:Species: Vibrio cholerae  
 C:Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004  
 C:Accession: B82223  
 R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
 Chardon, D.; Esmailova, M.D.; Vamathevan, J.; Baas, S.; Qin, H.; Dragol, I.; Sellers, F.  
 I., R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
 Nature 406, 477-483, 2000  
 A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
 A:Reference number: A82035; MUID:20406833; PMID:10952301  
 A:Accession: B82223  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-143 <HEI>  
 A:Cross-references: UNIPROT:Q9SKJ3; GB:AE004204; GB:AE003852; NID:G9655729; PIDN:AAF9441  
 A:Experimental source: serogroup O1, strain N16961, biotype El Tor  
 C:Genetics:  
 A:Gene: VC1253  
 A:Map position: 1

Query Match  
 Best Local Similarity 81.0%; Score 34; DB 2; Length 142;  
 Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFER 8  
 :|||  
 Db 86 KLPQFRR 93

RESULT 9  
 T36642  
 hypothetical protein SCH35.17 - Streptomyces coelicolor  
 C:Species: Streptomyces coelicolor  
 C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C:Accession: T36642  
 R:Olliver, K.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Raftery, M.A.  
 submitted to the EMBL Data Library, June 1999  
 A:Reference number: Z21610  
 A:Accession: T36642  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-189 <OLI>  
 A:Cross-references: UNIPROT:Q9X8Y9; EMBL:AL078610; PIDN:CA84410.1; GSPDB:GN00070; SCODE  
 A:Experimental source: strain A3(2)  
 C:Genetics:  
 A:Gene: SCODEB:SCH35.17

Query Match  
 Best Local Similarity 81.0%; Score 34; DB 2; Length 189;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFER 7  
 :|||  
 Db 105 KHPQPE 111

RESULT 10  
 AD2533  
 hypothetical protein alr7551 [imported] - Nostoc sp. (strain PCC 7120) plasmid pCC7120b  
 C:Species: Nostoc sp. PCC 7120  
 A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
 C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
 C:Accession: AD2533  
 R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kurita, T.; Sasamoto, S.; Watanabe, A.; Iriuch  
 Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 8, 205-213, 2001  
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Ana  
 A:Reference number: AB1807; MUID:21595285; PMID:11759840  
 A:Accession: AD2533  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-213 <KUR>  
 A:Cross-references: UNIPROT:Q8ZSF9; GB:AP003602; PIDN:BA877194.1; PID:G17134636; GSPDB:  
 A:Experimental source: strain PCC 7120  
 C:Genetics:  
 A:Gene: alr7551  
 A:Genome: plasmid

Query Match  
 Best Local Similarity 81.0%; Score 34; DB 2; Length 213;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFER 7  
 :|||  
 Db 114 RWHPOFQ 120

RESULT 11  
 H69297  
 conserved hypothetical protein AF0384 - Archaeoglobus fulgidus  
 C:Species: Archaeoglobus fulgidus  
 C:Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004  
 C:Accession: H69297  
 R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson  
 .; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.  
 Gload, A.; Zhou, L.; Overbeek, R.; Goeyne, J.D.; Weidman, J.F.; McDonald, L.  
 Nature 390, 364-370, 1997  
 A:Authors: Uetzerback, T.; Cotton, M.D.; Spriggs, T.; Artach, P.; Kaine, B.P.; Sykes, S  
 Smith, H.O.; Moese, C.R.; Venter, J.C.  
 A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeo  
 A:Reference number: A69250; MUID:98049343; PMID:9389475  
 A:Accession: H69297  
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A:Molecule type: DNA  
 A:Residues: 1-313 <KLE>  
 A:Cross-references: UNIPROT:Q29863; GB:AE001078; GB:AE000782; NID:G2689401; PIDN:AAB9085

Query Match  
 Best Local Similarity 81.0%; Score 34; DB 2; Length 313;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8  
 :|||  
 Db 267 YHPOFER 273

RESULT 12  
 T24938  
 hypothetical protein T15H9.1 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
 C:Accession: T24938  
 R:Gardner, A.  
 submitted to the EMBL Data Library, January 1995

A:Reference number: Z19958  
A:Accession: T24938  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1355 <MTL>  
A:Cross-references: UNIPROT:Q10005; EMBL:Z47356; PIDN:CAA87414.1; GSPDB:GN00020; CESP:TI  
C:Genetics:  
A:Experimental source: clone T15H9  
A:Gene: CESP:T15H9.1  
A:Map position: 2  
A:introns: 42/3; 74/3; 129/3; 151/3; 208/3; 249/2; 290/3  
C:Superfamily: heat shock protein dna; dna amino-terminal homology <DNT>  
F:4-89/Domain: dna amino-terminal homology <DNT>

Query Match 81.0%; Score 34; DB 2; Length 355;  
Best Local Similarity 62.5%; Pred. No. 58;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFER 8  
DB 245 QKHPRFR 252

RESULT 13  
T18140  
hypothetical protein A638R - Chlorella virus PBCV-1  
C:Species: Chlorella virus PBCV-1  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T18140  
R:Graves, M.V.; Van Etten, J.L.  
Submitted to the EMBL Data Library, May 1999  
A:Reference number: Z18806  
A:Accession: T18140  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-360 <GRA>  
A:Cross-references: UNIPROT:Q41120; EMBL:U42580; NID:94028896; PIDN:AAC96965.1  
A:Experimental source: specific host Chlorella strain NC64A  
C:Genetics:  
A>Note: A638R  
C:Superfamily: uncharacterized conserved protein SP0921

Query Match 81.0%; Score 34; DB 2; Length 360;  
Best Local Similarity 71.4%; Pred. No. 59;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8  
DB 233 KHPYER 239

RESULT 14  
C75580  
adenine deaminase-related protein - Deinococcus radiodurans (strain R1)  
C:Species: Deinococcus radiodurans  
C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
C:Accession: C75580  
R:White, O.; Eisen, J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;  
M.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.; Ma  
Science 286, 1571-1577, 1999  
A:Title: Genome sequence of the radioresistant bacterium Deinococcus radiodurans R1.  
A:Reference number: A75250; MUID:20036896; PMID:10567266  
A:Accession: C75580  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-376 <MHT>  
A:Cross-references: UNIPROT:Q9RYP2; GB:AE001863; GB:AE001825; NID:96460670; PIDN:AAF1237  
C:Genetics:  
A:Experimental source: strain R1  
A:Gene: DRA0268  
A:Map position: 2

Query Match 81.0%; Score 34; DB 2; Length 376;  
Best Local Similarity 62.5%; Pred. No. 62;  
Matches 5; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFER 8  
DB 315 RHHPFRSR 322

RESULT 15  
C82138  
conserved hypothetical protein VC1931 [imported] - Vibrio cholerae (strain N16961 serogr  
C:Species: Vibrio cholerae  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004  
C:Accession: C82138  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;  
Chadson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoti, I.; Sellers, P  
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833; PMID:10952301  
A:Accession: C82138  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-508 <HEI>  
A:Cross-references: UNIPROT:Q9KQR7; GB:AE004269; GB:AE003852; NID:93656466; PIDN:AAF9507  
C:Genetics:  
A:Gene: VC1931  
A:Map position: 1  
C:Superfamily: hypothetical protein b1706

Query Match 81.0%; Score 34; DB 2; Length 508;  
Best Local Similarity 71.4%; Pred. No. 85;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8  
DB 484 EHPRFR 490

Search completed: March 2, 2005, 12:28:55  
Job time: 7.11382 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 29.7236 Seconds

(without alignments)  
137.824 Million cell updates/sec

Title: SEQ9  
Perfect score: 42  
Sequence: 1 rxhpgfer 8

Scoring table: BLOSUM62X  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot\_03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	92.9	313	2	Q9P127
2	39	92.9	360	1	MDH_PYRHO
3	37	88.1	300	1	YB00_DEIRA
4	37	88.1	362	1	MDH_PYRAB
5	37	88.1	376	2	Q7X930
6	37	88.1	384	2	Q7V620
7	37	88.1	384	2	Q7WEC2
8	37	88.1	557	2	Q8XRVO
9	37	88.1	986	2	Q99FKO
10	35	83.3	129	2	Q9PAK0
11	35	83.3	221	2	Q83C26
12	35	83.3	328	2	Q6AC78
13	35	83.3	338	2	Q76230
14	35	83.3	338	2	Q981X8
15	35	83.3	420	2	Q42530
16	35	83.3	720	2	Q7R785
17	35	83.3	749	2	Q9V8J3
18	35	83.3	2713	2	Q6D9L1
19	35	83.3	2747	2	Q643C7
20	35	83.3	10820	2	Q7SPK2
21	34	81.0	26	2	Q92G31
22	34	81.0	55	2	Q92G36
23	34	81.0	77	2	Q86NS2
24	34	81.0	123	2	Q9MET4
25	34	81.0	142	2	Q9SKX3
26	34	81.0	147	2	Q8PAR9
27	34	81.0	167	2	Q89B59
28	34	81.0	182	2	Q88XG7
29	34	81.0	189	2	Q9X8Y9
30	34	81.0	193	2	Q6R266
31	34	81.0	207	2	Q6X127

32	34	81.0	213	2	Q8ZSF9	Q8ZSF9 anabaena sp
33	34	81.0	240	2	Q6TBT6	Q6TBT6 unclutred
34	34	81.0	313	1	Y384_ARCFU	Y384_ARCFU
35	34	81.0	335	2	Q7V101	Q7V101 prochloroc
36	34	81.0	342	2	Q8TKF9	Q8TKF9 methanosaic
37	34	81.0	344	2	Q6K9A8	Q6K9A8 oryza sativ
38	34	81.0	359	2	Q41120	Q41120 paramecium
39	34	81.0	366	2	Q9W083	Q9W083 drosophila
40	34	81.0	376	2	Q9RYP2	Q9RYP2 delnoccocus
41	34	81.0	388	2	Q96UG0	Q96UG0 metarhizium
42	34	81.0	396	2	Q726V6	Q726V6 desulfocidr
43	34	81.0	400	2	Q83A51	Q83A51 coxiella bu
44	34	81.0	418	2	Q9M554	Q9M554 euphorbia e
45	34	81.0	439	1	IDG1_DROVA	IDG1_DROVA drosophila

## ALIGNMENTS

RESULT 1					
ID	Q9P127	PRELIMINARY;	PRT;	313 AA.	
AC	Q9P127;				
DT	01-OCT-2000 (TREMBLrel. 15, Created)				
DT	01-OCT-2000 (TREMBLrel. 15, Last sequence update)				
DT	05-JUL-2004 (TREMBLrel. 27, Last annotation update)				
DE	HOM-TRF-85 tumor antigen (D0237H2.1).				
GN	Name=D0237H2.1;				
OS	Homo sapiens (Human).				
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;				
OC	Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.				
OX	NCBI_Taxid=9606;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	TISSUE=Testis;				
RX	MEDLINE=2028805; PubMed=12032826; DOI=10.1038/sj.onc.1205481;				
RA	Tureci O., Sahin U., Koslowski M., Bues B., Bell C., Ballweber P.,				
RA	Zwick C., Eberle T., Zuber M., Villena-Heinsen C., Seitz G.,				
RA	Pfeundschuh M.;				
RT	"A novel tumour associated leucine zipper protein targeting to sites				
RT	of gene transcription and splicing.";				
RL	Oncogene 21:3879-3888 (2002).				
RN	[2]				
RP	SEQUENCE FROM N.A.				
RA	Wilson S.;				
RL	Submitted (SEP-2000) to the EMBL/GenBank/DBJ databases.				
DR	EMBL; AF124430; AAF28870.1; -				
DR	EMBL; AL109751; CAC09922.1; -				
DR	SEQUENCE 313 AA; 35937 MW; E041911D9BA1DC8B CRC64;				
Query Match					
Best Local Similarity 92.9%; Score 39; DB 2; Length 313;					
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;					
QY	1	1	1	1	1
DB	182	182	182	182	182
RESULT 2					
ID	MDH_PYRHO	STANDARD;	PRT;	360 AA.	
AC	Q59028;				
DT	30-MAY-2000 (Rel. 39, Created)				
DT	30-MAY-2000 (Rel. 39, Last sequence update)				
DT	05-JUL-2004 (Rel. 44, Last annotation update)				
DE	Malate dehydrogenase (EC 1.1.1.37).				
GN	Name=mdh; OrderedAccession=PH1277;				
OS	Pyrococcus horikoshii.				
OC	Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;				
OC	Pyrococcus.				
OX	NCBI_Taxid=53953;				
RN	[1]				

```

RP SEQUENCE FROM N.A.
RC STRAIN=OT3.
RX MEDLINE=9834437; PubMed=9679194;
RA Kawarabayashi Y., Sawada M., Horikawa H., Halkawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Horiyama A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamura M., Ohkuchi Y.,
RA Funahashi T., Tanaka T., Kudoh K., Yamazaki J., Kishida N., Oguchi A.,
RA Aoki K.-I., Yoshizawa T., Nakamura Y., Robb F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.,
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.",
RL DNA Res. 5:155-76(1998).
CC -1- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the LDH2/MDH2 oxidoreductase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AP000005; BAA30380.1; -.
DR PIR; B71073; B71073.
DR InterPro; IPR003767; 1dh_2.
DR Pfam; PF0261b; 1dh_2; 1.
KM Complete proteome; NAD; Oxidoreductase; Tricarboxylic acid cycle.
SQ SEQUENCE 360 AA; 39751 MW; 6E9DB1E6CDE6E6F CRC64;

Query Match
Best Local Similarity 92.9%; Score 39; DB 1; Length 360;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFE 8
DB 311 RXHPOFE 318

RESULT 3
YB00 DEIRA
ID YB00 DEIRA STANDARD; PRT; 300 AA.
AC 09RVG9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 25-OCT-2004 (Rel. 45, Last sequence update)
DE Hypothetical UPF0176 protein DR1100.
GN OrderedLocuNames=DR1100;
OS Deinococcus radiodurans.
OC Bacteria; Deinococcus-Thermus; Deinococci; Deinococcales;
OC NCBI_TaxID=1299;
OX [1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=R1 / ATCC 13939 / DSM 20539 / NCTB 9279;
RX MEDLINE=2003696; PubMed=10567266; DOI=10.1126/science.286.5444.1571;
RA White O., Eison J.A., Haidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J., Lam P., McDonald L.A., Uitterlind T.R., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S.L., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome sequence of the radioresistant bacterium Deinococcus
RT radiodurans R1.",
RL Science 286:1571-1577(1999).
CC -1- SIMILARITY: Belongs to the UPF0176 family.
CC -1- SIMILARITY: Contains 1 rhodanese domain.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE001960; AAF10674.1; -.
DR PIR; G75436; G75436.
DR TIGR; DR1100; -.
DR HAMAP; MF_00469; -.
DR InterPro; IPR001792; Acylphosphatase.
DR InterPro; IPR001763; Rhodanese-like.
DR Pfam; PF00581; Rhodanese; 1.
DR SMART; SM00450; RHOD; 1.
DR PROSITE; PS02026; RHODANESF 3; 1.
KM Complete proteome; Hypothetical protein.
FT DOMAIN 128 222 Rhodanese.
SQ SEQUENCE 300 AA; 33238 MW; BE3D28BCDC43804 CRC64;

Query Match
Best Local Similarity 88.1%; Score 37; DB 1; Length 300;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFE 7
DB 261 RAHPQFE 267

RESULT 4
MDH_PYRAB
ID MDH_PYRAB STANDARD; PRT; 362 AA.
AC 09V0D5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 05-JUN-2004 (Rel. 44, Last annotation update)
DE Malate dehydrogenase (EC 1.1.1.37).
GN Name=mdh; OrderedLocuNames=PYRAB08550; ORFNames=PB1791;
OS Pyrococcus abyssi.
OC Archaea; Euryarchaeota; Thermococci; Thermococcales; Thermococcaceae;
OC Pyrococcus.
OC NCBI_TaxID=29292;
OX [1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=GES / Orsay;
RX MEDLINE=2511545; PubMed=12622808;
RA Cohen G.N., Barde V., Flament D., Galperin M., Hellig R., Lecompte O.,
RA Poch O., Prieur D., Querellou J., Ripp R., Thierry J.-C.,
RA Van der Oost J., Weissbach J., Zivanovic Y., Forterre P.;
RT "An integrated analysis of the genome of the hyperthermophilic
RT archaeon Pyrococcus abyssi.",
RL Mol. Microbiol. 47:1495-1512(2003).
CC -1- CATALYTIC ACTIVITY: (S)-malate + NAD(+) = oxaloacetate + NADH.
CC -1- SUBUNIT: Homodimer (By similarity).
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (Probable).
CC -1- SIMILARITY: Belongs to the LDH2/MDH2 oxidoreductase family.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AJ248285; CAB49765.1; -.
DR PIR; H75131; H75131.
DR InterPro; IPR003767; 1dh_2.
DR Pfam; PF02615; 1dh_2; 1.
KM Complete proteome; NAD; Oxidoreductase; Tricarboxylic acid cycle.
SQ SEQUENCE 362 AA; 39907 MW; 1E15C2B9E6BA5012 CRC64;

Query Match
Best Local Similarity 88.1%; Score 37; DB 1; Length 362;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
Qy 1 RXHPOFER 8
Db 311 RKGPDFER 318

RESULT 5
Q7X930 PRELIMINARY; PRT; 376 AA.
ID Q7X930
AC Q7X930;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
RT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
GN S-locus F-box protein 3.
OS Prunus avium (Cherry).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC Eucosids I; Rosales; Rosaceae; Amygdaloideae; Prunus.
OX NCBI_TaxId=42229;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=22763733; PubMed=12881505;
RA Yamane H., Ikeda K., Ushijima K., Sasaba H., Tao R.;
RT "A pollen-expressed gene for a novel protein with an F-box motif that
RT is very tightly linked to a gene for S-RNase in two species of cherry,
RT Prunus ceratua and P. avium."
RL Plant Cell Physiol. 44:764-769(2003).
DR EMBL; AB096857; BAC81148.1; -.
DR InterPro; IPR001810; F-box.
DR InterPro; IPR005627; F-box_assoc_1.
DR InterPro; IPR011043; Gal_oxid_central.
DR InterPro; IPR005054; RNA_rec_mot.
DR Pfam; PF00646; F-box; 1.
DR SMART; SM00256; FBOX; 1.
DR TIGRfam; TIGR01640; F-box_assoc_1; 1.
DR PROSITE; PS00030; RRM_RNP_1; UNKNOWN 1.
SQ SEQUENCE 376 AA; 43962 MW; 6076938579366587 CRC64;

Query Match
Best Local Similarity 88.1%; Score 37; DB 2; Length 376;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 2 XHPOFER 8
Db 63 HHPQFER 69

RESULT 6
Q7VS20 PRELIMINARY; PRT; 384 AA.
ID Q7VS20
AC Q7VS20;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
RT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
GN Hypothetical protein.
OS OrderedLocustNames=BP0636;
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxId=520;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Tohama I / ATCC BAA-589 / NCTC 13251;
RA MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA Parthill J., Sebalhia M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdeno-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Felwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,

```

```

RA Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica."
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640412; CAE44962.1; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 384 AA; 39424 MW; 4A23966B197193F2 CRC64;

Query Match
Best Local Similarity 88.1%; Score 37; DB 2; Length 384;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 2 XHPOFER 8
Db 130 GHPOFER 136

RESULT 7
Q7WEC2 PRELIMINARY; PRT; 384 AA.
ID Q7WEC2
AC Q7WEC2;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
RT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
GN Hypothetical protein.
OS OrderedLocustNames=BA4712;
OC Bordetella bronchiseptica (Alcaligenes bronchisepticus).
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Alcaligenaceae; Bordetella.
OX NCBI_TaxId=518;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RB50 / ATCC BAA-588;
RA MEDLINE=22827954; PubMed=12910271; DOI=10.1038/ng1227;
RA Parthill J., Sebalhia M., Preston A., Murphy L.D., Thomson N.R.,
RA Harris D.E., Holden M.T.G., Churcher C.M., Bentley S.D., Mungall K.L.,
RA Cerdeno-Tarraga A.-M., Temple L., James K.D., Harris B., Quail M.A.,
RA Achman M., Atkin R., Baker S., Basham D., Bason N., Cherevach I.,
RA Chillingworth T., Collins M., Cronin A., Davis P., Doggett J.,
RA Felwell T., Goble A., Hamlin N., Hauser H., Holtroyd S., Jagels K.,
RA Leather S., Moule S., Norberczak H., O'Neill S., Ormond D., Price C.,
RA Rabinowitch E., Rutter S., Sanders M., Saunders D., Seeger K.,
RA Sharp S., Simmonds M., Skelton J., Squares R., Squares S., Stevens K.,
RA Unwin L., Whitehead S., Barrell B.G., Maskell D.J.
RT "Comparative analysis of the genome sequences of Bordetella pertussis,
RT Bordetella parapertussis and Bordetella bronchiseptica."
RL Nat. Genet. 35:32-40(2003).
DR EMBL; BX640451; CAE35075.1; -.
KW Complete proteome; Hypothetical protein.
SQ SEQUENCE 384 AA; 39467 MW; 779AEC7C42AB9038 CRC64;

Query Match
Best Local Similarity 88.1%; Score 37; DB 2; Length 384;
Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
Qy 2 XHPOFER 8
Db 130 GHPOFER 136

RESULT 8
Q8XRVO PRELIMINARY; PRT; 557 AA.
ID Q8XRVO
AC Q8XRVO;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
RT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE PROBABLE TREHALOSE-6-PHOSPHATE SYNTHASE (ALPHA, ALPHA-TREHALOSE-
GN Name=RS01697; OrderedLocustNames=RS0731;
OS Ralstonia solanacearum (Pseudomonas solanacearum).

```



```

OX NCBI_TaxId=777;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=Nine Mile phase I / RSA 493;
RX MEDLINE=22608657; PubMed=12704232; DOI=10.1073/pnas.0931379100;
RA Sehnadri R., Paulsen I.T., Eisen J.A., Read T.D., Nelson K.E.,
RA Nelson W.C., Ward N.L., Tetteil H., Davidsen T.M., Beanan M.J.,
RA DeBoy R.T., Daugherty S.C., Brinkac L.M., Madupu R., Dodson R.J.,
RA Khouli H.M., Lee K.H., Carty H.A., Scanlan D., Heitzen R.A.,
RA Thompson H.A., Samuel J.E., Fraser C.M., Heidelberg J.F.,
RT "Complete genome sequence of the Q-fever pathogen, Coxiella
RT burnetii."
RL Proc. Natl. Acad. Sci. U.S.A. 100:5455-5460(2003).
DR EMBL: AE016964; AAC090814.1; -.
KM TIGR: CBU308; -.
KW Complete proteome.
SQ SEQUENCE 221 AA; 25813 MW; 87DB00A7B91578DF CRC64;

Query Match 83.3%; Score 35; DB 2; Length 221;
Best Local Similarity 75.0%; Pred. No. 86;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RXHPOFER 8
Db 168 RGHPRFER 175

RESULT 12
ID 06AC78 PRELIMINARY; PRT; 328 AA.
AC 06AC78;
DT 25-OCT-2004 (TReMBLrel. 28, Created)
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)
DE DnaJ protein..
GN Name=dnaJ; OrderedLocustNames=Lxxx23760;
OS Leifsonia xyli (subsp. xyli).
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;
OC Micrococcales; Microbacteriaceae; Leifsonia.
OK NCBI_TaxId=59736;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CTCB07;
RX PubMed=15305603;
RA Monteiro-Vitorello C.B., Camargo L.E.A., Van Sluys M.A.,
RA Kitajima J.P., Truffi D., do Amaral A.M., Harakava R.,
RA de Oliveira J.C.F., Wood D., de Oliveira M.C., Miyaki C.Y.,
RA Taktak M.A., da Silva A.C.R., Furlan L.R., Carrero D.M., Camarotte G.,
RA Almeida N.F., Jr., Carrier H., Coutinho L.L., El-Dorry H.A.,
RA Ferro M.I.T., Gagliardi P.R., Gigliotti E., Goldman M.H.S.,
RA Goldman G.H., Kimura E.T., Ferro E.S., Kuramae E.E., Lemos E.G.M.,
RA Lemos M.V.F., Mauro S.M.Z., Machado M.A., Marino C.L., Menck C.F.,
RA Nunes L.R., Oliveira R.C., Pereira G.G., Siqueira W., de Souza A.A.,
RA Teat S.M., Zanca A.S., Simpson A.J.G., Brumley S.M., Seubal J.C.;
RT "The genome sequence of the Gram-positive sugarcane pathogen Leifsonia
RT xyli subsp. xyli."
RL Mol. Plant Microbe Interact. 17:827-836(2004).
DR EMBL: AE016822; AAT90014.1; -.
DR GO: GO:0051082; F:unfolding protein binding; IEA.
DR GO: GO:0006457; P:protein folding; IEA.
DR InterPro: IPR002939; DnaJ_C.
DR InterPro: IPR001623; DnaJ_N.
DR InterPro: IPR008971; HSP40_DnaJ_pap.
DR InterPro: IPR003095; Hsp_DnaJ.
DR Pfam: PF00226; DnaJ_1.
DR Pfam: PF01556; DnaJ_C_1.
DR PRINTS: PR00625; DnaJPROTEIN.
DR SMART: SM00271; DnaJ_1.
DR PROSITE: PS00636; DnaJ_1; 1.
DR PROSITE: PS50076; DnaJ_2; 1.
KW Chaperone; Complete proteome.
SQ SEQUENCE 328 AA; 34507 MW; 21596D5320173639 CRC64;

```

```

Query Match 83.3%; Score 35; DB 2; Length 328;
Best Local Similarity 75.0%; Pred. No. 1.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 RXHPOFER 8
Db 217 RGHPRFER 224

RESULT 13
ID 076230 PRELIMINARY; PRT; 338 AA.
AC 076230;
DT 01-NOV-1998 (TReMBLrel. 08, Created)
DT 01-NOV-1998 (TReMBLrel. 08, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Chaperone.
DE Chaperone.
GN Name=DnaJ;
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OK NCBI_TaxId=5653;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CL;
RX MEDLINE=98418771; PubMed=9747975; DOI=10.1016/S0166-6851(98)00080-2;
RA Brindaud F., Vedremme C., Cuivillier A., Parry D., Baltz D., Tetaud E.,
RA Pays E., Venegas J., Merlin G., Baltz T.,
RT "Conserved organization of genes in trypanosomatids."
RL Mol. Biochem. Parasitol. 94:249-264(1998).
DR EMBL: AF031927; AAC32777.1; -.
DR PIR: T30538; T30538.
DR HSP: P25685; 1HDJ.
DR GO: GO:0051082; F:unfolding protein binding; IEA.
DR GO: GO:0006457; P:protein folding; IEA.
DR InterPro: IPR002939; DnaJ_C.
DR InterPro: IPR001623; DnaJ_N.
DR InterPro: IPR008971; HSP40_DnaJ_pap.
DR InterPro: IPR003095; Hsp_DnaJ.
DR Pfam: PF00226; DnaJ_1.
DR Pfam: PF01556; DnaJ_C_1.
DR PRINTS: PR00625; DnaJPROTEIN.
DR SMART: SM00271; DnaJ_1.
DR PROSITE: PS00636; DnaJ_1; 1.
DR PROSITE: PS50076; DnaJ_2; 1.
KW Chaperone.
SQ SEQUENCE 338 AA; 36535 MW; 95BA7EA791E2A19A CRC64;

Query Match 83.3%; Score 35; DB 2; Length 338;
Best Local Similarity 62.5%; Pred. No. 1.4e+02;
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RXHPOFER 8
Db 242 RGHPRFER 249

RESULT 14
ID 09B1X8 PRELIMINARY; PRT; 338 AA.
AC 09B1X8;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)
DE Co-chaperone protein.
GN Name=DnaJ-like;
OS Trypanosoma cruzi.
OC Eukaryota; Euklenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.
OK NCBI_TaxId=5653;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21570179; PubMed=11551903; DOI=10.1074/jbc.M102427200;
RA Salmon D., Monteiro-Lomeij M., Goldenberg S.;
RT "A DnaJ-like protein homologous to the yeast co-chaperone Sisl (Tcd6p)

```

RT is involved in initiation of translation in *Trypanosoma cruzi*."  
RL J. Biol. Chem. 276:43970-43979(2001).  
DR EMBL; AF345336; AK19734.1; -.  
DR HSPB, P25685; IHDJ.  
DR GO; GO:0051082; F:unfolded protein binding; IEA.  
DR GO; GO:0006457; P:protein folding; IEA.  
DR InterPro; IPR002939; DnaJ\_C.  
DR InterPro; IPR001623; DnaJ\_N.  
DR InterPro; IPR008971; HSP40\_DnaJ\_pep.  
DR InterPro; IPR003095; Hsp\_DnaJ.  
DR Pfam; PF00226; DnaJ\_1.  
DR Pfam; PF01556; DnaJ\_C; 1.  
DR PRINTS; PR00625; DnaJPROTEIN.  
DR SMART; SM00271; DnaJ; 1.  
DR PROSITE; PS00636; DnaJ\_1; 1.  
DR PROSITE; PS0076; DnaJ\_2; 1.  
DR KW Chaperone.  
SQ SEQUENCE 338 AA; 36561 MW; 90A9E4B04508E655 CRC64;

Query Match 83.3%; Score 35; DB 2; Length 338;  
Best Local Similarity 62.5%; Pred. NO. 1.4e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RXHPQFER 8  
::|||:  
Db 242 KHPHPRFR 249

RESULT 15  
Q42530 PRELIMINARY; PRT; 420 AA.  
AC Q42530;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
DE DnaJ homolog.  
DE Name=atj;  
GN Arabidopsis thaliana (Mouse-ear cress).  
OS Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosid1; Brassicales; Brassicaceae; Arabidopsids.  
OC NCBI\_TaxID=3702;  
OX [1]  
RN SEQUENCE FROM N.A.  
RP MEDLINE=20027448; PubMed=10557255;  
RX Zhou R., Kroczyńska B., Mierzyk J.A.,  
RT "ACU3 (Accession No. U22340), an Arabidopsis thaliana J-Protein  
Homologous to Saccharomyces cerevisiae YDulp. (PGR99-162)."  
RL Plant Physiol. 121:1053-1054(1999).  
DR EMBL; U22340; AAB49030.1; -.  
DR PIR; S71199; S71199.  
DR HSPB, P25685; IHDJ.  
DR GO; GO:0051082; F:unfolded protein binding; IEA.  
DR GO; GO:0006457; P:protein folding; IEA.  
DR InterPro; IPR002939; DnaJ\_C.  
DR InterPro; IPR001305; DnaJ\_CXXCXG.  
DR InterPro; IPR001623; DnaJ\_N.  
DR InterPro; IPR008971; HSP40\_DnaJ\_pep.  
DR InterPro; IPR003095; Hsp\_DnaJ.  
DR Pfam; PF00226; DnaJ\_1.  
DR Pfam; PF01556; DnaJ\_C; 1.  
DR Pfam; PF00684; DnaJ\_CXXCXG; 1.  
DR PRINTS; PR00625; DnaJPROTEIN.  
DR SMART; SM00271; DnaJ; 1.  
DR PROSITE; PS00636; DnaJ\_1; 1.  
DR PROSITE; PS0076; DnaJ\_2; 1.  
DR PROSITE; PS00637; DnaJ\_CXXCXG; 1.  
KW Chaperone; Repeat.  
SQ SEQUENCE 420 AA; 46444 MW; C4C12848F61AD445 CRC64;

Query Match 83.3%; Score 35; DB 2; Length 420;  
Best Local Similarity 62.5%; Pred. NO. 1.8e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

Qy 1 RXHPQFER 8  
::|||:  
Db 260 KHPHPRFR 267

Search completed: March 2, 2005, 12:44:20  
Job time : 30.7236 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 36.0325 Seconds  
(without alignments)  
85.869 Million cell updates/sec

Title: SEQ9

Perfect score: 42  
Sequence: 1 rxhpgfer 8

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*

1: geneseqp1980a:.\*  
2: geneseqp1990a:.\*  
3: geneseqp2000a:.\*  
4: geneseqp2001a:.\*  
5: geneseqp2002a:.\*  
6: geneseqp2003a:.\*  
7: geneseqp2003b:.\*  
8: geneseqp2004a:.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	92.9	231	7	ADB65422
2	39	92.9	313	4	AAB75608
3	39	92.9	359	8	AD841924
4	39	92.9	360	8	ADN18635
5	37	88.1	11	2	AAR52694
6	37	88.1	11	2	AAR52693
7	37	88.1	12	2	AAR52697
8	37	88.1	362	8	ADS43180
9	36	85.7	482	7	ADG95241
10	36	85.7	619	6	ABU19450
11	35	83.3	418	6	ABP80316
12	35	83.3	749	4	ABR62997
13	35	83.3	2747	7	ADH86071
14	34	81.0	8	2	AAR59212
15	34	81.0	8	4	AAB35433
16	34	81.0	8	4	AAB35442
17	34	81.0	8	4	AAB35442
18	34	81.0	8	5	AAB68616
19	34	81.0	8	5	ABO19065
20	34	81.0	8	5	ABR57464
21	34	81.0	8	5	ABR74486
22	34	81.0	8	6	ABG73584
23	34	81.0	8	6	AAE37230
24	34	81.0	8	6	ABP60361
25	34	81.0	8	6	ABP60368
				7	ADA09808

26	34	81.0	8	7	ADB84588	ADB84588
27	34	81.0	8	7	ADB85500	ADB85500
28	34	81.0	8	7	ADD29930	ADD29930
29	34	81.0	8	8	AD006951	AD006951
30	34	81.0	8	8	ADP90832	ADP90832
31	34	81.0	8	8	ADP90832	ADP90832
32	34	81.0	8	8	ADP90832	ADP90832
33	34	81.0	8	8	ADP90832	ADP90832
34	34	81.0	8	8	ADP90832	ADP90832
35	34	81.0	10	4	AA979662	AA979662
36	34	81.0	10	4	AA979662	AA979662
37	34	81.0	10	5	AA980475	AA980475
38	34	81.0	10	6	ABP56623	ABP56623
39	34	81.0	10	8	ADP11064	ADP11064
40	34	81.0	10	8	ADP11064	ADP11064
41	34	81.0	10	8	ADP11064	ADP11064
42	34	81.0	11	6	AA838373	AA838373
43	34	81.0	11	8	ADP20228	ADP20228
44	34	81.0	13	2	AA878379	AA878379
45	34	81.0	19	6	ABG74881	ABG74881
				6	ABG74882	ABG74882

#### ALIGNMENTS

RESULT 1  
ADB65422  
ID ADB65422 standard; protein, 231 AA.

AC ADB65422;

DT 04-DEC-2003 (first entry)

DE Human protein encoded by clone TESIT120180600.

XX Human; pharmaceutical; diagnostic; gene therapy; tissue regeneration;

XX cell regeneration; membrane protein; signal transduction-related protein;

XX transcription-related protein; osteoporosis; neurological disease;

XX cancer; tumour.

XX Homo sapiens.

XX EP1308459-A2.

XX 07-MAY-2003.

XX 28-MAR-2002; 2002EP-00007401.

XX 05-NOV-2001; 2001JP-00379298.

XX 25-JAN-2002; 2002US-00350978.

XX (HELI-) HELIX RES INST.

XX (REAS-) RES ASSOC BIOTECHNOLOGY.

XX Isegai T, Suiyama T, Otsuki T, Wakamatsu A, Sato H, Ishii S;

XX Yamamoto U, Isono Y, Hio Y, Otsuka K, Nagai R, Tamechika I;

XX Seki N, Yoshikawa T, Otsuka M, Nagahari K, Masuno Y;

XX WPI; 2003-450961/43.

XX N-PSDB; ADB63452.

XX New polynucleotides and polypeptides, useful for developing a diagnostic

XX marker or medicines for regulation of their expression and activity, or

XX as targets of gene therapy.

Claim 1, Page; 222pp; English.

The invention discloses a polynucleotide comprising a sequence selected  
from 1970 Fully defined nucleotide sequences which encode novel  
polypeptides. Also claimed is a polypeptide encoded by the polynucleotide  
or its partial peptide, an antibody binding to the polypeptide or peptide  
of the polynucleotide, immunologically assaying the polypeptide or  
peptide of the polynucleotide by contacting the polypeptide or peptide

CC with the antibody of the encoded protein, and observing the binding  
CC between the two, a transformant carrying the polynucleotide in an  
CC expressible manner and an antisense polynucleotide. The oligonucleotide  
CC is useful as a primer for synthesizing the polynucleotide, or as a probe  
CC for detecting the polynucleotide. The polynucleotide and encoded  
CC proteins are useful as pharmaceutical agents and many disease-related  
CC genes may be included in them, for developing a diagnostic marker or  
CC medicines for regulation of their expression and activity, or as targets  
CC of gene therapy. The genes are involved in tissue and/or cell  
CC regeneration. Membrane proteins, signal transduction-related proteins,  
CC transcription-related proteins, disease-related proteins and genes  
CC encoding them can be used as indicators for diseases (e.g. osteoporosis,  
CC neurological diseases, cancer, tumors). The CDNA may be used to regulate  
CC the activity or expression of the encoded protein to treat diseases. The  
CC sequence presented is a protein of the invention. Note: Some of the  
CC sequence data for this patent is not represented in the printed  
CC European Patent Office.

SQ Sequence 231 AA;

Query Match 92.9%; Score 39; DB 7; Length 231;

Best Local Similarity 75.0%; Pred. No. 41;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOPER 8

Db 100 RHHPOPER 107

RESULT 2

ID AAB75608 standard; protein; 313 AA.

XX AAB75608;

DT 10-APR-2001 (first entry)

XX Human cancer associated antigen precursor HOM-TES-85 SEQ ID NO:11.

XX Human; cancer associated antigen precursor; cancer associated antigen;

XX seminoma; HLA; human leukocyte antigen; cytostatic; gene therapy;

XX vaccine; cancer.

XX Homo sapiens.

XX WO200100874-A2.

XX 04-JAN-2001.

XX 23-JUN-2000; 2000WO-US017207.

XX 30-JUN-1999; 99US-00346498.

XX (LUDW-) LUDWIG INST CANCER RES.

XX Sahin U, Tureci O, Pfreundschuh M;

XX WPI; 2001-112465/12.

XX N-PSDB; AAF26957.

XX Diagnosing a disorder characterized by expression of a human cancer

XX associated antigen precursor, comprises detecting interaction of an agent

XX with a nucleic acid molecule encoding the antigen precursor.

XX Example 2; Page 97; 126pp; English.

XX The present invention describes a method for diagnosing a disorder

XX characterized by expression of a human cancer associated antigen (CAA)

XX precursor (I) coded by a NA Group 1 nucleic acid molecule (N1) comprising

XX contacting the biological sample with an agent (A) that specifically

XX binds to N1, (I) or its fragment, complexed with an human leukocyte

XX antigen (HLA) molecule and determining the interaction between the agent

CC and N1 or (I). (I) has cytostatic activity and can be used in gene  
CC therapy and vaccine production. The method can be used for treating a  
CC subject with a condition characterized by expression of (I) in cells of a  
CC subject. AAB75607 and AAB75608 represent proteins from human cancer  
CC associated antigen precursors, and AAB75609 to AAB75802 represent HLA  
CC class I binding motifs in human cancer associated antigen precursors  
CC given in the exemplification of the present invention

SQ Sequence 313 AA;

Query Match 92.9%; Score 39; DB 4; Length 313;

Best Local Similarity 75.0%; Pred. No. 57;

Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOPER 8

Db 182 RHHPOPER 189

RESULT 3

ID ADS41924 standard; protein; 359 AA.

XX ADS41924;

DT 02-DEC-2004 (first entry)

XX Bacterial polypeptide #20354.

XX Recombinant DNA construct; transformed plant; improved plant property;

XX cold tolerance; heat tolerance; drought tolerance; herbicide; osmosis;

XX pathogen tolerance; pest tolerance; plant disease resistance;

XX cell cycle pathway modification; plant growth regulator;

XX homologous recombination; seed oil yield; protein yield; carbohydrate;

XX nitrogen; phosphorus; photosynthesis; lignin; galactomannan;

XX bacterial polypeptide.

XX Bacteria.

XX US2003233675-A1.

XX 18-DEC-2003.

XX 20-FEB-2003; 2003US-00369493.

XX 21-FEB-2002; 2002US-0360039P.

XX (CAOY/) CAO Y.

XX (HINKLE/) HINKLE G J.

XX (SLAT/) SLATER S C.

XX (CHEN/) CHEN X.

XX (GOLD/) GOLDMAN B S.

XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;

XX WPI; 2004-061375/06.

XX New recombinant DNA construct comprising a promoter positioned to provide

XX for expression of a polynucleotide encoding a polypeptide from a

XX microbial source, useful for producing plants with improved properties.

XX Claim 1; SEQ ID NO 20354; 122pp; English.

XX The invention relates to a recombinant DNA construct comprising a

XX promoter functional in a plant cell, where the promoter is positioned to

XX provide for expression of a polynucleotide encoding a polypeptide from a

XX microbial source. The invention also relates to a transformed plant

XX comprising the recombinant DNA construct and a method of producing a

XX transformed plant having an improved property. The plant is a crop plant

XX such as maize or soybean. The method of producing a transformed plant

XX having an improved property comprises transforming a plant with the

XX recombinant DNA construct and growing the transformed plant, where the

XX polynucleotide or polypeptide is useful for improving plant properties.



CC The recombinant DNA construct is useful for producing plants with  
CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
CC increased resistance to plant disease, better growth rate by modification  
CC of the cell cycle pathway with plant growth regulators, increased rate of  
CC homologous recombination, modified seed oil or protein yield and/or  
CC content, improved yield by modification of carbohydrate, nitrogen or  
CC phosphorus use and/or uptake, by modification of photosynthesis or by  
CC providing improved plant growth and development under at least one stress  
CC condition, improved lignin production or improved galactomanan  
CC production. This sequence represents a bacterial polypeptide used in the  
CC scope of the invention. Note: The sequence data for this patent did not  
CC form part of the printed specification but was obtained in electronic  
CC format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
SQ Sequence 359 AA;  
Query Match 92.9%; Score 39; DB 8; Length 359;  
Best Local Similarity 75.0%; Pred. No. 66;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 RXHPOFER 8  
Db 311 RKHPEFER 318  
RESULT 4  
ADN18635 ADN18635 standard; protein; 360 AA.  
XX  
XX ADN18635; 1  
XX  
XX 02-DEC-2004 (first entry)  
DT  
XX Bacterial polypeptide #1288.  
DE  
XX Recombinant DNA construct; transformed plant; improved plant property;  
XX cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic;  
XX pathogen tolerance; pest tolerance; plant disease resistance;  
XX cell cycle pathway modification; plant growth regulator;  
XX homologous recombination; seed oil yield; protein yield; carbohydrate;  
XX nitrogen; phosphorus; photosynthesis; lignin; galactomanan;  
XX bacterial polypeptide.  
XX  
XX Bacteria.  
OS  
XX US2003233675-A1.  
PN  
XX 18-DEC-2003.  
PD  
XX 20-FEB-2003; 2003US-00369493.  
PF  
XX 21-FEB-2002; 2002US-0360039P.  
PR  
XX (CAOY/) CAO Y.  
XX (HINK/) HINKLE G J.  
XX (SLAT/) SLATER S C.  
XX (CHEN/) CHEN X.  
XX (GOLD/) GOLDMAN B S.  
XX  
XX Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS.  
XX WPI; 2004-061375/06.  
XX  
XX New recombinant DNA construct comprising a promoter positioned to provide  
XX for expression of a polynucleotide encoding a polypeptide from a  
XX microbial source, useful for producing plants with improved properties.  
XX  
XX Claim 1; SEQ ID NO 1288; 122bp; English.  
XX  
XX The invention relates to a recombinant DNA construct comprising a  
XX promoter functional in a plant cell, where the promoter is positioned to  
XX provide for expression of a polynucleotide encoding a polypeptide from a

CC microbial source. The invention also relates to a transformed plant  
CC comprising the recombinant DNA construct and a method of producing a  
CC transformed plant having an improved property. The plant is a crop plant  
CC such as maize or soybean. The method of producing a transformed plant  
CC having an improved property comprises transforming a plant with the  
CC recombinant DNA construct and growing the transformed plant, where the  
CC polynucleotide or polypeptide is useful for improving plant properties.  
CC The recombinant DNA construct is useful for producing plants with  
CC improved plant properties, e.g. improved cold, heat or drought tolerance,  
CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,  
CC increased resistance to plant disease, better growth rate by modification  
CC of the cell cycle pathway with plant growth regulators, increased rate of  
CC homologous recombination, modified seed oil or protein yield and/or  
CC content, improved yield by modification of carbohydrate, nitrogen or  
CC phosphorus use and/or uptake, by modification of photosynthesis or by  
CC providing improved plant growth and development under at least one stress  
CC condition, improved lignin production or improved galactomanan  
CC production. This sequence represents a bacterial polypeptide used in the  
CC scope of the invention. Note: The sequence data for this patent did not  
CC form part of the printed specification but was obtained in electronic  
CC format from USPTO at [seqdata.uspto.gov/sequence.html](http://seqdata.uspto.gov/sequence.html).  
SQ Sequence 360 AA;  
Query Match 92.9%; Score 39; DB 8; Length 360;  
Best Local Similarity 75.0%; Pred. No. 66;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 RXHPOFER 8  
Db 311 RKHPEFER 318  
RESULT 5  
AAR52694 AAR52694 standard; protein; 11 AA.  
ID  
XX AAR52694;  
AC  
XX 10-JAN-1995 (first entry)  
DT  
XX PASK46-p14KH encoded C-terminal streptavidin-binding sequence.  
DE  
XX  
XX Streptavidin binding peptide; fusion protein; immunoglobulin; Ig;  
XX heavy chain variable region; affinity chromatography; purification;  
XX peptide tag.  
XX  
XX Synthetic.  
OS  
XX  
XX Key Location/Qualifiers  
FT Peptide 4..11  
FT /note= "streptavidin-binding octapeptide fused to C-terminus of VH chain"  
FT  
XX  
XX GB2272698-A.  
XX  
XX 25-MAY-1994.  
PD  
XX 01-NOV-1993; 93GB-00022501.  
PF  
XX 03-NOV-1992; 92DE-04237113.  
PR  
XX (BIOA-) INST BIOANALYTIK GEMEINNUTZIGE GMBH.  
XX  
XX Skerra A, Schmidt T;  
XX  
XX WPI: 1994-153484/19.  
XX  
XX N-PSDB; AAO62669.  
XX  
XX New fusion peptide(s) - have easily controlled binding properties and are  
XX capable of binding to streptavidin.  
XX  
XX Disclosure; Page 11; 53pp; English.  
XX  
XX

XX Six derivatives of pASK46 (a plasmid for expression of the D1.3Fv  
 CC fragment in E.coli) were produced which encode 4 different peptides at  
 CC the C-terminus of VH and 2 peptides at the C-terminus of VL of the D1.3Fv  
 CC fragment. The peptides fused to the C-terminus are all examples of  
 CC streptavidin-binding peptides corresponding to a generic formula (see  
 CC AAR52698). The peptides do not interfere with the protein function but  
 CC facilitate purification by conferring streptavidin-binding properties on  
 CC the fusion protein

SQ Sequence 11 AA;

Query Match 88.1%; Score 37; DB 2; Length 11;  
 Best Local Similarity 85.7%; Pred. No. 3.7;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFER 8  
 :|||||  
 DB 5 LHPQFER 11

RESULT 6  
 AAR52693  
 ID AAR52693 standard; protein; 11 AA.

AC AAR52693;  
 XX  
 DT 10-JAN-1995 (first entry)

DE pASK46-p11XH encoded C-terminal streptavidin-binding sequence.

KM Streptavidin binding peptide; fusion protein; immunoglobulin; Ig;  
 KM heavy chain variable region; affinity chromatography; purification;  
 KM peptide tag.

OS Synthetic.

Key Location/Qualifiers  
 FH 4. 11  
 FT /note= "streptavidin-binding octapeptide fused to C-  
 FT terminus of VH chain"

GB2272698-A.

25-MAY-1994.

01-NOV-1993; 93GB-00022501.

03-NOV-1992; 92DE-04237113.

(BIOA-) INST BIOANALYTIK GEMEINNUTZIGE GMBH.

Skerra A, Schmidt T;

WPI, 1994-153484/19.

DR N-PSDB; AAQ62668.

PT New fusion peptide(s) - have easily controlled binding properties and are  
 PT capable of binding to streptavidin.

PS Disclosure; Page 11; 53pp; English.

XX Six derivatives of pASK46 (a plasmid for expression of the D1.3Fv  
 CC fragment in E.coli) were produced which encode 4 different peptides at  
 CC the C-terminus of VH and 2 peptides at the C-terminus of VL of the D1.3Fv  
 CC fragment. The peptides fused to the C-terminus are all examples of  
 CC streptavidin-binding peptides corresponding to a generic formula (see  
 CC AAR52698). The peptides do not interfere with the protein function but  
 CC facilitate purification by conferring streptavidin-binding properties on  
 CC the fusion protein

SQ Sequence 11 AA;

Query Match 88.1%; Score 37; DB 2; Length 11;  
 Best Local Similarity 85.7%; Pred. No. 3.7;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFER 8  
 :|||||  
 DB 5 RHPQFER 11

RESULT 7  
 AAR52697  
 ID AAR52697 standard; protein; 12 AA.

AC AAR52697;

DT 10-JAN-1995 (first entry)

DE pASK46-p11XL encoded C-terminal streptavidin-binding sequence.

KM Streptavidin binding peptide; fusion protein; immunoglobulin; Ig;  
 KM light chain variable region; affinity chromatography; purification;  
 KM peptide tag.

OS Synthetic.

Key Location/Qualifiers  
 FH 5. 12  
 FT /note= "streptavidin-binding octapeptide fused to C-  
 FT terminus of VL chain"

GB2272698-A.

25-MAY-1994.

01-NOV-1993; 93GB-00022501.

03-NOV-1992; 92DE-04237113.

(BIOA-) INST BIOANALYTIK GEMEINNUTZIGE GMBH.

Skerra A, Schmidt T;

WPI, 1994-153484/19.

DR N-PSDB; AAQ62672.

PT New fusion peptide(s) - have easily controlled binding properties and are  
 PT capable of binding to streptavidin.

PS Disclosure; Page 11; 53pp; English.

XX Six derivatives of pASK46 (a plasmid for expression of the D1.3Fv  
 CC fragment in E.coli) were produced which encode 4 different peptides at  
 CC the C-terminus of VH and 2 peptides at the C-terminus of VL of the D1.3Fv  
 CC fragment. The peptides fused to the C-terminus are all examples of  
 CC streptavidin-binding peptides corresponding to a generic formula (see  
 CC AAR52698). The peptides do not interfere with the protein function but  
 CC facilitate purification by conferring streptavidin-binding properties on  
 CC the fusion protein

SQ Sequence 12 AA;

Query Match 88.1%; Score 37; DB 2; Length 12;  
 Best Local Similarity 85.7%; Pred. No. 4;  
 Matches 6; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFER 8  
 :|||||  
 DB 6 RHPQFER 12

RESULT 8  
 ADS43180  
 ID ADS43180 standard; protein; 362 AA.

```
XX AC ADS43180;
XX DT 02-DEC-2004 (first entry)
XX DE Bacterial polypeptide #21610.
XX KM Recombinant DNA construct; transformed plant; improved plant property;
XX KM cold tolerance; heat tolerance; drought tolerance; herbicide; osmotic;
XX KM pathogen tolerance; pest tolerance; plant disease resistance;
XX KM cell cycle pathway modification; plant growth regulator;
XX KM homologous recombination; seed oil yield; protein yield; carbohydrate;
XX KM nitrogen; phosphorus; photosynthesis; lignin; galactomannan;
XX KM bacterial polypeptide.
XX OS Bacteria.
XX PN US2003233675-A1.
XX PD 18-DEC-2003.
XX PF 20-FEB-2003; 2003US-00369493.
XX PR 21-FEB-2002; 2002US-0360039P.
XX PA (CAO/) CAO Y.
XX PA (HINK/) HINKLE G J.
XX PA (SLAT/) SLATER S C.
XX PA (CHEN/) CHEN X.
XX PA (GOLD/) GOLDMAN B S.
XX PI Cao Y, Hinkle GJ, Slater SC, Chen X, Goldman BS;
XX DR WPI; 2004-061375/06.
XX PT New recombinant DNA construct comprising a promoter positioned to provide
XX PT for expression of a polynucleotide encoding a polypeptide from a
XX PT microbial source, useful for producing plants with improved properties.
XX PS Claim 1; SEQ ID NO 21610; 122pp; English.
XX CC The invention relates to a recombinant DNA construct comprising a
XX CC promoter functional in a plant cell, where the promoter is positioned to
XX CC provide for expression of a polynucleotide encoding a polypeptide from a
XX CC microbial source. The invention also relates to a transformed plant
XX CC comprising the recombinant DNA construct and a method of producing a
XX CC transformed plant having an improved property. The plant is a crop plant
XX CC such as maize or soybean. The method of producing a transformed plant
XX CC having an improved property comprises transforming a plant with the
XX CC recombinant DNA construct and growing the transformed plant, where the
XX CC polynucleotide or polypeptide is useful for improving plant properties.
XX CC The recombinant DNA construct is useful for producing plants with
XX CC improved plant properties, e.g. improved cold, heat or drought tolerance,
XX CC tolerance to herbicides, extreme osmotic conditions, pathogens or pests,
XX CC increased resistance to plant disease, better growth rate by modification
XX CC of the cell cycle pathway with plant growth regulators, increased rate of
XX CC homologous recombination, modified seed oil or protein yield and/or
XX CC content, improved yield by modification of carbohydrate, nitrogen or
XX CC phosphorus use and/or uptake, by modification of photosynthesis or by
XX CC providing improved plant growth and development under at least one stress
XX CC condition, improved lignin production or improved galactomannan
XX CC production. This sequence represents a bacterial polypeptide used in the
XX CC scope of the invention. Note: The sequence data for this patent did not
XX CC form part of the printed specification but was obtained in electronic
XX CC format from USPTO at seqdata.uspto.gov/sequence.html.
XX SQ Sequence 362 AA;
XX QY Query Match 88.1%; Score 37; DB 8; Length 362;
XX QY Best Local Similarity 75.0%; Pred. No. 1.6e+02;
XX QY Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
DB 311 RXHPFER 318
XX AC AD95241
XX ID ADC95241 standard; protein; 482 AA.
XX AC ADC95241;
XX DT 01-JAN-2004 (first entry)
XX DE E. faecium protein sequence SEQ ID 4868.
XX KM Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;
XX KM abdominal-pelvic infection.
XX OS Enterococcus faecium.
XX PN US6583275-B1.
XX PD 24-JUN-2003.
XX PF 30-JUN-1998; 98US-00107532.
XX PR 02-JUL-1997; 97US-0051571P.
XX PR 14-MAY-1998; 98US-0085598P.
XX PA (GENO-) GENOME THERAPEUTICS CORP.
XX PI Doucette-Stamm LA, Bush D;
XX DR WPI; 2003-799836/75.
XX DR N-PSDB; ADC91587.
XX PT New isolated nucleic acid derived from Enterococcus faecium encoding an
XX PT Enterococcus faecium polypeptide useful for detection, prevention and
XX PT treatment of a pathological condition resulting from a bacterial
XX PT infection.
XX PS Example 1; SEQ ID NO 4868; 243pp; English.
XX CC The invention relates to an isolated nucleic acid derived from
XX CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having
XX CC one of 10 fully defined sequences given in the (or comprising 40
XX CC sequential nucleotides chosen from any of the nucleic acids, its
XX CC complement or sequences hybridising to it). Also included are a
XX CC recombinant vector comprising the nucleic acid operably linked to
XX CC transcription regulatory element, a cell comprising the vector and a
XX CC single-stranded probe comprising the nucleic acid. The nucleic acids are
XX CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.
XX CC The nucleic acids are useful for diagnosing pathological conditions
XX CC resulting from E. faecium bacterial infection (e.g. urinary tract
XX CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic
XX CC infection) and for screening drugs such as agonists and antagonists. The
XX CC nucleic acid is useful for recombinant production of Candida albicans -
XX CC derived peptides or antisense polypeptides. Pharmaceutical compositions
XX CC and vaccines containing the nucleic acid are useful for preventing or
XX CC treating Enterococcus faecium infections. The present sequence represents
XX CC one if the disclosed E. faecium proteins.
XX SQ Sequence 482 AA;
XX QY Query Match 85.7%; Score 36; DB 7; Length 482;
XX QY Best Local Similarity 62.5%; Pred. No. 3.3e+02;
XX QY Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;
```

RESULT 10

ABU19450  
ID ABU19450 standard; protein; 619 AA.  
XX  
AC ABU19450;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by prokaryotic essential gene #4977.  
XX  
KM Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Borrelia cepacia.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
XX  
PR 06-SEP-2001; 2001US-00948993.  
XX  
PR 25-OCT-2001; 2001US-0342923P.  
XX  
PR 08-FEB-2002; 2002US-00072851.  
XX  
PR 06-MAR-2002; 2002US-0362699F.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
FI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zykkind JM;  
FI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
XX WPI; 2003-029926/02.  
XX  
DR N-PSDB; ACP23320.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 47374; 1766bp; English.  
XX  
XX The invention relates to an isolated nucleic acid comprising any one of  
XX the 6213 antisense sequences given in the specification where expression  
XX of the nucleic acid inhibits proliferation of a cell. Also included are:  
XX (1) a vector comprising a promoter operably linked to the nucleic acid  
XX encoding a polypeptide whose expression is inhibited by the antisense  
XX nucleic acid; (2) a host cell containing the vector; (3) an isolated  
XX polypeptide or its fragment whose expression is inhibited by the  
XX antisense nucleic acid; (4) an antibody capable of specifically binding  
XX the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
XX proliferation; (7) identifying a compound that influences the activity of  
XX the gene product or that has an activity against a biological pathway  
XX required for proliferation, or that inhibits cellular proliferation; (8)  
XX identifying a gene required for cellular proliferation or its gene product  
XX pathway in which a proliferation-required gene or its gene product lies  
XX or a gene on which the test compound that inhibits proliferation of an  
XX organism acts; (9) manufacturing an antibiotic; (10) profiling a  
XX compound's activity; (11) a culture comprising strains in which the gene  
XX product is overexpressed or underexpressed; (12) determining the extent  
XX to which each of the strains is present in a culture or collection of  
XX strains; or (13) identifying the target of a compound that inhibits the  
XX proliferation of an organism. The antisense nucleic acids are useful for  
XX identifying proteins or screening for homologous nucleic acids required  
XX for cellular proliferation to isolate candidate molecules for rational  
XX drug discovery programs, or for screening homologous nucleic acids  
XX required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
XX *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
XX the target prokaryotic essential genes. Note: The sequence data for this  
XX patent did not form part of the printed specification, but was obtained  
XX in electronic format directly from WIPO at  
XX ftp.wipo.int/pub/published\_pct\_sequences  
XX  
XX Sequence 619 AA;

Query Match 85.7%; Score 36; DB 6; Length 619;  
Best Local Similarity 75.0%; Pred. No. 4.3e+02;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
OY 1 RXHPQFER 8  
DB 339 RHFOIER 346  
RESULT 11  
ID ABP80316 standard; protein; 418 AA.  
XX  
AC ABP80316;  
XX  
DT 07-MAR-2003 (first entry)  
XX  
DE N. gonorrhoeae amino acid sequence SEQ ID 7162.  
XX  
KM Antibacterial; infection; vaccine; gene therapy.  
XX  
OS Neisseria gonorrhoeae.  
XX  
PN WO200279243-A2.  
XX  
PD 10-OCT-2002.  
XX  
PF 12-FEB-2002; 2002WO-IB002069.  
XX  
PR 12-FEB-2001; 2001GB-00003424.  
XX  
XX (CHIR-) CHIRON SPA.  
XX  
FI Fontana MR, Piza M, Maignani V, Monaci E;  
XX  
XX WPI; 2003-058415/05.  
XX  
DR N-PSDB; AB241286.  
XX  
PT New protein from Neisseria gonorrhoeae, useful for the manufacture of a  
PT medicament for treating or preventing N. gonorrhoeae infection.  
XX  
PS Disclosure; Page 708; 815bp; English.  
XX  
XX The present invention relates to proteins from Neisseria gonorrhoeae.  
XX Also disclosed are the nucleic acid molecules encoding the proteins and  
XX antibodies that specifically bind to the proteins. The composition  
XX comprising the protein, nucleic acid or antibody is useful for the  
XX manufacture of a medicament for treating or preventing N. gonorrhoeae  
XX infection, this may be in the form of a vaccine or gene therapy.  
XX Sequences given in records ABP76736-ABP81046 represent nucleic acid  
XX molecules of the invention  
XX  
XX Sequence 418 AA;  
XX  
OY Query Match 83.3%; Score 35; DB 6; Length 418;  
DB Best Local Similarity 62.5%; Pred. No. 4.3e+02;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;  
OY 1 RXHPQFER 8  
DB 16 RHPRDR 23  
RESULT 12  
ID ABP82997 standard; protein; 749 AA.  
XX  
AC ABP82997;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
XX Drosophila melanogaster polypeptide SEQ ID NO 15783.  
XX  
XX

KW Drosophila; developmental biology; cell signalling; insecticide;  
 XX pharmaceutical.  
 OS Drosophila melanogaster.  
 XX WO200171042-A2.  
 XX 27-SEP-2001.  
 XX  
 PF 23-MAR-2001; 2001WO-US009231.  
 XX  
 XX 23-MAR-2000; 2000US-0191637P.  
 PR 11-JUL-2000; 2000US-00614150.  
 XX  
 PA (PEKE ) PE CORP NY.  
 XX  
 PI Venter JC, Adams M, Li PWD, Myers EW;  
 DR WPI; 2001-656860/75.  
 DR N-PSDB; ABL07100.  
 XX  
 XX New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signalling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 15783; 21bp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (AB16176-AB130511), expressed DNA  
 CC sequences (AB101840-AB16175) and the encoded proteins (AAB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 749 AA;  
 Query Match 83.3%; Score 35; DB 4; Length 749;  
 Best Local Similarity 75.0%; Pred. No. 8.1e+02;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 RXHPOPER 8  
 DB 546 RSHPOFVR 553  
 RESULT 13  
 ADE86071  
 ID ADE86071 standard; protein; 2747 AA.  
 XX  
 AC ADE86071;  
 XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Streptomyces hygroscopicus non-ribosomal peptide synthetase mppA.  
 XX  
 KW Peptide synthetase; enzyme; antibiotic; mppA; AC98.  
 XX  
 OS Streptomyces hygroscopicus.  
 XX  
 PN WO2003082909-A1.  
 XX  
 PD 09-OCT-2003.  
 XX  
 PF 28-MAR-2003; 2003WO-US009746.  
 XX  
 PR 29-MAR-2002; 2002US-0368713P.  
 XX  
 PA (AMHP ) WYETH.  
 XX

PI Hucul JA, Macgarvey N, Greenstein M;  
 XX WPI; 2003-812536/76.  
 DR N-PSDB; ADE86070.  
 XX  
 PT New non-ribosomal peptide synthetase, useful for preparing antibacterial  
 PT peptides, derived from Streptomyces, also related nucleic acid and  
 PT modified antibiotics.  
 XX  
 PS Claim 10; SEQ ID NO 2; 163bp; English.  
 XX  
 CC The present sequence is the protein sequence of mppA, a component of the  
 CC non-ribosomal peptide synthetase (NRPS) from Streptomyces hygroscopicus  
 CC NS17 (NRRL 30439) that is responsible for the production of cyclic  
 CC peptide antibiotic AC98. The NRPS complex exists as 2 separate  
 CC components, mppA and mppB ADE86072, both involved in the synthesis of the  
 CC core AC98. MppA is composed of 3 minimal modules, where each module is  
 CC comprised of an adenylation, thiolation and condensation domain. MppA  
 CC conjugates a serine to a glycine to produce a peptide. The peptide is  
 CC then conjugated (through the glycine) to a phenylalanine. Each amino acid  
 CC is incorporated into the peptide chain by a unique module. After addition  
 CC of the phenylalanine, the peptide chain is then transferred to the mppB  
 CC component. MppB is about 295 kDa. The invention provides a method of  
 CC producing NRPS using transformed host cells. It also provides a method of  
 CC producing a cyclic peptide antibiotic, such as AC98, using the NRPS  
 CC comprised of mppA and mppB. The antibiotic is preferably a  
 CC lipoglycopeptide antibiotic with activity against Gram-positive  
 CC pathogens. Also provided are methods of modifying NRPS, e.g. by  
 CC replacement of an adenylation domain, to produce an antibiotic having a  
 CC modified peptide core, and a method for evaluating the structural regions  
 CC of the modified peptide.  
 XX  
 SQ Sequence 2747 AA;  
 Query Match 83.3%; Score 35; DB 7; Length 2747;  
 Best Local Similarity 75.0%; Pred. No. 3.3e+03;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 1 RXHPOPER 8  
 DB 899 RDHPOVER 906  
 RESULT 14  
 AAM59212  
 ID AAM59212 standard; peptide; 8 AA.  
 XX  
 AC AAM59212;  
 XX  
 DT 27-AUG-1998 (first entry)  
 XX  
 DE Streptavidin tagged peptide ligand #2.  
 XX  
 KW Streptavidin; ligand; binding affinity; mutant; isolation; purification;  
 KW recover; immobilise.  
 XX  
 OS Synthetic.  
 XX  
 PN EP835934-A2.  
 XX  
 PD 15-APR-1998.  
 XX  
 PF 09-OCT-1997; 97EP-00117504.  
 XX  
 PR 10-OCT-1996; 96DE-01041876.  
 XX  
 PA (BIOA-) INST BIOANALYTIK GMBH.  
 XX  
 PI Skerra A, Voss S;  
 XX  
 DR WPI; 1998-218868/20.  
 XX  
 PT Streptavidin mutants with higher binding affinity for peptide ligands -

PT have mutation in amino acid region 44-53, used to isolate, purify or  
 PT determine fusion proteins including these ligands.

PS Claim 10, Page 11; 21pp; German.

XX AAW59211 and AAW59212 are ligands used in a method to assay binding  
 CC affinity of streptavidin mutants. These mutants have a mutation within  
 CC the amino acid (aa) region 44-53 of the wild-type protein show a higher  
 CC binding affinity than the wild-type for peptide ligands that include the  
 CC sequence of formula  $\text{Trp-X-His-Pro-Gln-Phe-Y-Z}$  where X = any aa; Y and Z  
 CC are both Gly, or Y = Glu and Z = Arg or Lys. Recombinant streptavidin  
 CC mutants can be used to isolate, purify and determine proteins or to  
 CC determine/recover substances that contain streptavidin-binding groups.  
 CC Such compounds may also be used to immobilise fusions on microtitre  
 CC plates, microbeads or sensor chips

XX SQ Sequence 8 AA;

Query Match 81.0%; Score 34; DB 2; Length 8;  
 Best Local Similarity 71.4%; Pred. No. 1.8e+06;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFER 8  
 :|||||:  
 2 SHPQFEK 8

RESULT 15  
 AAB35433  
 ID AAB35433 standard; peptide; 8 AA.

XX AAB35433;

DT 23-MAY-2001 (first entry)

XX Epitope peptide #3.

KW Nascent protein detection; protein analysis; aminoacylated tRNA;  
 KW BODIPY marker; disease diagnosis.

XX Unidentified.

XX WO200114578-A1.

PD 01-MAR-2001.

PF 23-AUG-2000; 2000WO-US023233.

PR 25-AUG-1999; 99US-00382736.

PR 25-AUG-1999; 99US-00382950.

PA (AMBE-) AMBERGEN INC.

PI Rochschild KJ, Gite S, Olejnik J;

DR WPI; 2001-168972/17.

PT Method for detecting nascent proteins by fluorescence comprises  
 PT misaminoacylating a tRNA molecule with a marker compound, useful for  
 PT detecting mutations in proteins, e.g. cancer.

PS Disclosure; Page 47; 204pp; English.

XX The present invention describes a method of detecting nascent proteins  
 CC involving aminoacylating a tRNA molecule with a 4,4-difluoro-4-bora-3a,4a  
 CC -diaz-a-s-indacene (BODIPY) marker leading to the production of a  
 CC misaminoacylated tRNA. This enables the detection, isolation and analysis  
 CC of nascent proteins using UV without the usual accompanying radioactivity  
 CC problems. It may be used to detect mutations, for example in cancer,  
 CC Duchenne muscular dystrophy, adenomatous polyposis coli and colon cancer

XX Sequence 8 AA;

Query Match 81.0%; Score 34; DB 4; Length 8;  
 Best Local Similarity 71.4%; Pred. No. 1.8e+06;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPQFER 8  
 :|||||:  
 DB 2 SHPQFEK 8

Search completed: March 2, 2005, 13:02:46  
 Job time : 36.0325 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005: Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 24.878 Seconds  
(without alignments)  
105.489 Million cell updates/sec

Title: SEQ9

Perfect score: 42  
Sequence: 1 rxhpofcr 8

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

- 1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*
- 17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep:\*
- 19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*
- 20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	39	92.9	231	15	US-10-104-047-3576
2	39	92.9	359	15	US-10-369-493-20354
3	39	92.9	360	15	US-10-369-493-1288
4	37	88.1	362	15	US-10-369-493-21610
5	36	85.7	619	15	US-10-282-122A-47374
6	35	83.3	36	14	US-10-026-578B-11
7	35	83.3	2747	15	US-10-402-842-2
8	34	81.0	8	9	US-09-809-517A-9
9	34	81.0	8	9	US-09-973-145-7
10	34	81.0	8	12	US-09-813-197-8
11	34	81.0	8	13	US-10-104-218-5
12	34	81.0	8	13	US-10-208-357-9
13	34	81.0	8	14	US-10-001-934-8

14	34	81.0	8	14	US-10-026-578B-2	Sequence 2, Appli
15	34	81.0	8	14	US-10-026-578B-9	Sequence 9, Appli
16	34	81.0	8	14	US-10-174-368A-7	Sequence 7, Appli
17	34	81.0	8	14	US-10-264-127-8	Sequence 8, Appli
18	34	81.0	8	15	US-10-339-712-8	Sequence 8, Appli
19	34	81.0	8	15	US-10-339-712-8	Sequence 67, Appli
20	34	81.0	8	15	US-10-275-046-4	Sequence 4, Appli
21	34	81.0	8	15	US-10-425-000-76	Sequence 76, Appli
22	34	81.0	8	15	US-10-424-999-25	Sequence 25, Appli
23	34	81.0	8	15	US-10-358-283-23	Sequence 23, Appli
24	34	81.0	8	16	US-10-628-432-11	Sequence 41, Appli
25	34	81.0	8	17	US-10-494-248-17	Sequence 17, Appli
26	34	81.0	8	17	US-10-634-645-11	Sequence 11, Appli
27	34	81.0	8	17	US-10-719-523-8	Sequence 8, Appli
28	34	81.0	9	9	US-09-983-067-3	Sequence 3, Appli
29	34	81.0	10	9	US-09-809-517A-6	Sequence 6, Appli
30	34	81.0	10	15	US-10-147-211A-20	Sequence 20, Appli
31	34	81.0	11	15	US-10-354-983-29	Sequence 29, Appli
32	34	81.0	11	16	US-10-628-432-25	Sequence 25, Appli
33	34	81.0	13	15	US-10-338-592-30	Sequence 30, Appli
34	34	81.0	21	9	US-09-809-517A-30	Sequence 30, Appli
35	34	81.0	22	9	US-09-809-517A-31	Sequence 33, Appli
36	34	81.0	24	9	US-09-809-517A-31	Sequence 31, Appli
37	34	81.0	24	14	US-10-026-578B-4	Sequence 3, Appli
38	34	81.0	24	14	US-10-026-578B-4	Sequence 4, Appli
39	34	81.0	25	9	US-09-809-517A-34	Sequence 34, Appli
40	34	81.0	36	14	US-10-026-578B-10	Sequence 10, Appli
41	34	81.0	117	10	US-09-977-137A-4	Sequence 4, Appli
42	34	81.0	117	10	US-09-977-137A-5	Sequence 5, Appli
43	34	81.0	117	10	US-09-977-137A-7	Sequence 7, Appli
44	34	81.0	117	10	US-09-977-137A-8	Sequence 8, Appli
45	34	81.0	117	10	US-09-977-137A-9	Sequence 9, Appli

## ALIGNMENTS

RESULT 1  
US-10-104-047-3576  
; Sequence 3576, Application US/10104047  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: HELIX RESEARCH INSTITUTE  
; TITLE OF INVENTION: No. US20030236392A1 full length cDNA  
; FILE REFERENCE: H1-A0105  
; CURRENT APPLICATION NUMBER: US/10/104,047  
; PRIOR APPLICATION NUMBER:  
; PRIOR FILING DATE:  
; NUMBER OF SEQ ID NOS: 4096  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 3576  
; TYPE: PRF  
; LENGTH: 231  
; ORGANISM: Homo sapiens  
US-10-104-047-3576

Query Match 92.9%; Score 39; DB 15; Length 231;  
Best Local Similarity 75.0%; Pred. No. 50;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPQFER 8  
|:|:|:|:  
Db 100 RXHPQFER 107

RESULT 2  
US-10-369-493-20354  
; Sequence 20354, Application US/10369493  
; Publication No. US20030236392A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.

```

; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20354
; LENGTH: 359
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-20354

Query Match          92.9%; Score 39; DB 15; Length 359;
Best Local Similarity 75.0%; Pred. No. 77;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 RXHPOFER 8
DB      311 RKHPFER 318

RESULT 3
US-10-369-493-1288
; Sequence 1288, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374.
; SEQ ID NO 1288
; LENGTH: 360
; TYPE: PRT
; ORGANISM: Pyrococcus horikoshii
US-10-369-493-1288

Query Match          92.9%; Score 39; DB 15; Length 360;
Best Local Similarity 75.0%; Pred. No. 77;
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY      1 RXHPOFER 8
DB      311 RKHPFER 318

RESULT 4
US-10-369-493-21610
; Sequence 21610, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493

```

```

; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 21610
; LENGTH: 362
; TYPE: PRT
; ORGANISM: Pyrococcus abyssi
US-10-369-493-21610

Query Match          88.1%; Score 37; DB 15; Length 362;
Best Local Similarity 75.0%; Pred. No. 1.7e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY      1 RXHPOFER 8
DB      311 RKHPFER 318

RESULT 5
US-10-282-122A-47374
; Sequence 47374, Application US/10282122A
; Publication No. US20040029129A1
; GENERAL INFORMATION:
; APPLICANT: Wang, Liangsu
; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haeselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zykkind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: EITRA 034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47374
; LENGTH: 619
; TYPE: PRT
; ORGANISM: Burkholderia cepacia
US-10-282-122A-47374

Query Match          85.7%; Score 36; DB 15; Length 619;
Best Local Similarity 75.0%; Pred. No. 4.4e+02;
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

```



QY 1 RXHPOFER 8  
|:|||||  
Db 339 RHPQFER 346.

RESULT 6  
US-10-026-578B-11  
; Sequence 11, Application US/10026578B  
; Publication No. US20030083474A1  
; GENERAL INFORMATION:  
; APPLICANT: IBA (GmbH)  
; APPLICANT: Schmidt, Thomas  
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Tag  
; FILE REFERENCE: 100810.01US1  
; CURRENT APPLICATION NUMBER: US/10/026,578B  
; CURRENT FILING DATE: 2002-11-11  
; PRIOR APPLICATION NUMBER: DE 101 13 776.1  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: PCT/EP01/11846  
; PRIOR FILING DATE: 2001-10-12  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 11  
; LENGTH: 36  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic Peptide  
; NAME/KEY: MISC FEATURE  
; LOCATION: (9)\_(128)  
; OTHER INFORMATION: Of the five (GGGS) repeats between position 9 and 28, one or more  
; OTHER INFORMATION: repeats may be missing. However, at least one (GGGS) repeat will  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Artificial Sequence represents peptide binding module  
US-10-026-578B-11

Query Match 83.3%; Score 35; DB 14; Length 36;  
Best Local Similarity 62.5%; Pred. No. 41;  
Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFER 8  
|:|||||  
Db 29 QSHPOFEK 36

RESULT 7  
US-10-402-842-2  
; Sequence 2, Application US/10402842  
; Publication No. US20030219872A1  
; GENERAL INFORMATION:  
; APPLICANT: Magarey, Nathan A.  
; APPLICANT: Hucul, John A.  
; TITLE OF INVENTION: NON-RIBOSOMAL PEPTIDE SYNTHETASES AND ASSOCIATED BIOSYNTHETIC GEN  
; FILE REFERENCE: 0630/17854-US1  
; CURRENT APPLICATION NUMBER: US/10/402,842  
; CURRENT FILING DATE: 2003-03-28  
; PRIOR APPLICATION NUMBER: US 60/368,713  
; PRIOR FILING DATE: 2002-03-30  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 2  
; LENGTH: 2747  
; TYPE: PRT  
; ORGANISM: Streptomyces hygroscopicus  
US-10-402-842-2

Query Match 83.3%; Score 35; DB 15; Length 2747;  
Best Local Similarity 75.0%; Pred. No. 2,8e+03;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFER 8  
|:|||||  
Db 899 RHPQFER 906

RESULT 8  
US-09-809-517A-9  
; Sequence 9, Application US/09809517A  
; Patent No. US20020034733A1  
; GENERAL INFORMATION:  
; APPLICANT: Lohning, Corinna  
; TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote  
; FILE REFERENCE: MORPHO/11  
; CURRENT APPLICATION NUMBER: US/09/809,517A  
; CURRENT FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: EP 99114072.4  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: EP 00103551.8  
; PRIOR FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 9  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module  
US-09-809-517A-9

Query Match 81.0%; Score 34; DB 9; Length 8;  
Best Local Similarity 71.4%; Pred. No. 1.2e+06;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8  
|:|||||  
Db 2 SHPOFEK 8

RESULT 9  
US-09-973-145-7  
; Sequence 7, Application US/09973145  
; Patent No. US20020132248A1  
; GENERAL INFORMATION:  
; APPLICANT: Rothschild, Kenneth J.  
; APPLICANT: Gite, Sedana  
; APPLICANT: Olejnik, Jerzy  
; TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins  
; FILE REFERENCE: AMBER-06819  
; CURRENT APPLICATION NUMBER: US/09/973,145  
; CURRENT FILING DATE: 2001-10-09  
; PRIOR APPLICATION NUMBER: 09/382,950  
; PRIOR FILING DATE: 1999-08-25  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 7  
; LENGTH: 8  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Synthetic  
US-09-973-145-7

Query Match 81.0%; Score 34; DB 9; Length 8;  
Best Local Similarity 71.4%; Pred. No. 1.2e+06;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8  
|:|||||  
Db 2 SHPOFEK 8

```
RESULT 10
US-09-813-197-8
; Sequence 8, Application US/09813197
; Publication No. US2005009013A1
; GENERAL INFORMATION:
; APPLICANT: Rothschild, Kenneth
; APPLICANT: Gite, Sadanand
; APPLICANT: Olejnik, Jerzy
; TITLE OF INVENTION: Method for the Detection, Analysis and Isolation of Nascent Prot
; FILE REFERENCE: AMBR-03951
; CURRENT APPLICATION NUMBER: US/09/813,197
; CURRENT FILING DATE: 2001-03-20
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic
US-09-813-197-8
```

```
Query Match      81.0%; Score 34; DB 12; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQPER 8
       :|||||:
Db      2 SHPOPEK 8
```

```
RESULT 11
US-10-104-218-5
; Sequence 5, Application US/10104218
; Publication No. US20020177196A1
; GENERAL INFORMATION:
; APPLICANT: MAIER, Thomas
; APPLICANT: GABBERT, Carsten
; TITLE OF INVENTION: PROCESS FOR PREPARING NON-PROTEINOGENIC L-AMINO ACIDS
; FILE REFERENCE: MAIER, T. ET AL.-2
; CURRENT APPLICATION NUMBER: US/10/104,218
; CURRENT FILING DATE: 2002-03-22
; PRIOR APPLICATION NUMBER: German No. US20020177196A1 101 16 881.0
; PRIOR FILING DATE: 2001-04-04
; PRIOR APPLICATION NUMBER: German No. US20020177196A1 101 21 515.0
; PRIOR FILING DATE: 2001-05-03
; NUMBER OF SEQ ID NOS: 5
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 5
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; NAME/KEY: PEPTIDE
; LOCATION: (1)..(8)
; OTHER INFORMATION: StreptagII affinity peptide for protein purification
US-10-104-218-5
```

```
Query Match      81.0%; Score 34; DB 13; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQPER 8
       :|||||:
Db      2 SHPOPEK 8
```

```
RESULT 12
US-10-208-357-9
; Sequence 9, Application US/10208357
; Publication No. US20020182687A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Kurz, Markus
; APPLICANT: Lohse, Peter
; APPLICANT: Wagner, Richard
; TITLE OF INVENTION: Peptide Acceptor Ligation Methods
; FILE REFERENCE: 50036/031002
; CURRENT APPLICATION NUMBER: US/10/208,357
; CURRENT FILING DATE: 2002-07-30
; PRIOR APPLICATION NUMBER: US/09/619,103
; PRIOR FILING DATE: 2000-07-19
; PRIOR APPLICATION NUMBER: 60/145,834
; PRIOR FILING DATE: 1999-07-27
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: designed sequence to act as an identifying tag
US-10-208-357-9
```

```
Query Match      81.0%; Score 34; DB 13; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQPER 8
       :|||||:
Db      2 SHPOPEK 8
```

```
RESULT 13
US-10-001-934-8
; Sequence 8, Application US/10001934
; Publication No. US20030032782A1
; GENERAL INFORMATION:
; APPLICANT: NAGY, ZOLTAN
; TITLE OF INVENTION: HUMAN PEPTIDES/PROTEINS CAUSING OR LEADING TO THE
; FILE REFERENCE: GRCG-P01-003
; CURRENT APPLICATION NUMBER: US/10/001,934
; CURRENT FILING DATE: 2001-11-15
; NUMBER OF SEQ ID NOS: 63
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 8
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-001-934-8
```

```
Query Match      81.0%; Score 34; DB 14; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      2 XHPQPER 8
       :|||||:
Db      2 SHPOPEK 8
```

```
RESULT 14
US-10-026-578B-2
; Sequence 2, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
```

```

; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 2
; LENGTH: 8
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; NAME/KEY: misc feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-2

```

```

Query Match      81.0%; Score 34; DB 14; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      2 XHPQFER 8
:|||||:
Db      2 SHPOFEK 8

```

```

RESULT 15
US-10-026-578B-9
; Sequence 9, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 9
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-9

```

```

Query Match      81.0%; Score 34; DB 14; Length 8;
Best Local Similarity 71.4%; Pred. No. 1.2e+06;
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy      2 XHPQFER 8
:|||||:
Db      2 SHPOFEK 8

```

Search completed: March 2, 2005, 14:18:51  
 Job time : 24.878 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 9.10569 Seconds  
(without alignments)  
65.585 Million cell updates/sec

Title: SEQ9  
Perfect score: 42  
Sequence: 1 rxhpgfer 8

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgn2\_6/pcodata/1/1aa/5A.COMB.pep:\*  
2: /cgn2\_6/pcodata/1/1aa/5B.COMB.pep:\*  
3: /cgn2\_6/pcodata/1/1aa/6A.COMB.pep:\*  
4: /cgn2\_6/pcodata/1/1aa/6B.COMB.pep:\*  
5: /cgn2\_6/pcodata/1/1aa/PCUS.COMB.pep:\*  
6: /cgn2\_6/pcodata/1/1aa/backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	36	85.7	482	4 US-09-107-532A-4868	Sequence 4868, Ap
2	35	83.3	128	4 US-09-270-767-32326	Sequence 32326, A
3	35	83.3	128	4 US-09-270-767-47543	Sequence 47543, A
4	34	81.0	8	3 US-08-948-097-2	Sequence 2, Appl
5	34	81.0	8	3 US-09-382-950-7	Sequence 7, Appl
6	34	81.0	8	3 US-09-382-736B-8	Sequence 8, Appl
7	34	81.0	8	4 US-09-619-103-9	Sequence 9, Appl
8	34	81.0	8	4 US-10-104-218-5	Sequence 5, Appl
9	34	81.0	8	4 US-09-809-517A-9	Sequence 9, Appl
10	34	81.0	10	4 US-09-809-517A-6	Sequence 6, Appl
11	34	81.0	13	2 US-08-687-559-30	Sequence 30, Appl
12	34	81.0	13	4 US-09-401-415-30	Sequence 30, Appl
13	34	81.0	21	4 US-09-809-517A-30	Sequence 30, Appl
14	34	81.0	22	4 US-09-809-517A-33	Sequence 33, Appl
15	34	81.0	24	4 US-09-809-517A-31	Sequence 31, Appl
16	34	81.0	25	4 US-09-809-517A-34	Sequence 34, Appl
17	34	81.0	117	4 US-09-977-137A-4	Sequence 4, Appl
18	34	81.0	117	4 US-09-977-137A-5	Sequence 5, Appl
19	34	81.0	117	4 US-09-977-137A-7	Sequence 7, Appl
20	34	81.0	117	4 US-09-977-137A-8	Sequence 8, Appl
21	34	81.0	117	4 US-09-977-137A-9	Sequence 9, Appl
22	34	81.0	117	4 US-09-977-137A-10	Sequence 10, Appl
23	34	81.0	117	4 US-09-977-137A-11	Sequence 11, Appl
24	34	81.0	117	4 US-09-977-137A-12	Sequence 12, Appl
25	34	81.0	118	4 US-09-977-137A-6	Sequence 6, Appl
26	34	81.0	275	4 US-09-270-767-58917	Sequence 58917, A
27	34	81.0	777	4 US-09-902-540-9844	Sequence 9844, Ap

28	34	81.0	801	4 US-09-270-767-43549	Sequence 43549, A
29	34	81.0	3666	2 US-08-222-617A-12	Sequence 12, Appl
30	34	81.0	3727	2 US-08-222-617A-27	Sequence 27, Appl
31	34	81.0	3778	2 US-08-222-617A-2	Sequence 2, Appl
32	33	78.6	103	4 US-09-270-767-41422	Sequence 41422, A
33	33	78.6	114	4 US-09-270-767-42719	Sequence 42719, A
34	33	78.6	205	4 US-09-252-991A-25397	Sequence 25397, A
35	33	78.6	369	4 US-09-252-991A-17585	Sequence 17585, A
36	33	78.6	413	4 US-09-723-546-11	Sequence 11, Appl
37	33	78.6	444	1 US-09-178-002-2	Sequence 2, Appl
38	33	78.6	466	3 US-08-704-711A-17	Sequence 17, Appl
39	33	78.6	466	3 US-09-521-220-17	Sequence 17, Appl
40	33	78.6	467	1 US-09-178-002-4	Sequence 4, Appl
41	33	78.6	467	3 US-09-391-104-24	Sequence 24, Appl
42	33	78.6	468	3 US-08-448-489-13	Sequence 13, Appl
43	33	78.6	468	3 US-09-689-730-13	Sequence 13, Appl
44	33	78.6	477	4 US-09-252-991A-16778	Sequence 16778, A
45	33	78.6	650	4 US-09-107-532A-5521	Sequence 5521, Ap

## ALIGNMENTS

```

RESULT 1
US-09-107-532A-4868
; Sequence 4868, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESS: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 4868:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 482 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHEICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (8) LOCATION 1...482
; SEQUENCE DESCRIPTION: SEQ ID NO: 4868:
US-09-107-532A-4868

```

Query Match 85.7%; Score 36; DB 4; Length 482;  
 Best Local Similarity 62.5%; Pred. No. 45;  
 Matches 5; Conservative 3; Mismatches 0; Indels 0; Gaps 0;

QY 1 RXHPOFER 8  
 DB 409 KKHPEFER 416

RESULT 2  
 US-09-270-767-32326  
 ; Sequence 32326, Application US/09270767

; Patent No. 6703491  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Homburger et al.  
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
 ; FILE REFERENCE: File Reference: 7326-094  
 ; CURRENT APPLICATION NUMBER: US/09/270,767  
 ; CURRENT FILING DATE: 1999-03-17  
 ; NUMBER OF SEQ ID NOS: 62517  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 32326  
 ; LENGTH: 128  
 ; TYPE: PRT

; ORGANISM: Drosophila melanogaster  
 US-09-270-767-32326

Query Match 83.3%; Score 35; DB 4; Length 128;  
 Best Local Similarity 75.0%; Pred. No. 16;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFER 8  
 DB 43 RSHPOFER 50

RESULT 3  
 US-09-270-767-47543

; Sequence 47543, Application US/09270767  
 ; Patent No. 6703491  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Homburger et al.  
 ; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
 ; FILE REFERENCE: File Reference: 7326-094  
 ; CURRENT APPLICATION NUMBER: US/09/270,767  
 ; CURRENT FILING DATE: 1999-03-17  
 ; NUMBER OF SEQ ID NOS: 62517  
 ; SOFTWARE: PatentIn Ver. 2.0  
 ; SEQ ID NO 47543  
 ; LENGTH: 128  
 ; TYPE: PRT

; ORGANISM: Drosophila melanogaster  
 US-09-270-767-47543

Query Match 83.3%; Score 35; DB 4; Length 128;  
 Best Local Similarity 75.0%; Pred. No. 16;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 RXHPOFER 8  
 DB 43 RSHPOFER 50

RESULT 4  
 US-08-948-097-2

; Sequence 2, Application US/08948097C  
 ; Patent No. 6103493  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Skeira, Arne  
 ; APPLICANT: Vos, Selma  
 ; TITLE OF INVENTION: Streptavidin Muteins  
 ; FILE REFERENCE: HUBR 1119

CURRENT APPLICATION NUMBER: US/08/948,097C  
 ; CURRENT FILING DATE: 1997-10-09  
 ; EARLIER APPLICATION NUMBER: DE 196 41 876.3  
 ; EARLIER FILING DATE: 1996-10-10  
 ; NUMBER OF SEQ ID NOS: 17  
 ; SEQ ID NO 2  
 ; LENGTH: 8  
 ; TYPE: PRT

; ORGANISM: Artificial sequence  
 ; FEATURE:  
 ; NAME/KEY: BINDING  
 ; OTHER INFORMATION: Binding ligand for streptavidin

US-08-948-097-2

Query Match 81.0%; Score 34; DB 3; Length 8;  
 Best Local Similarity 71.4%; Pred. No. 4.1e+05;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8  
 DB 2 SHPOFER 8

RESULT 5  
 US-09-382-950-7

; Sequence 7, Application US/09382950  
 ; Patent No. 6303337  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rothschild, Kenneth  
 ; APPLICANT: Gite, Sadanand  
 ; APPLICANT: Olejnik, Jerry  
 ; TITLE OF INVENTION: N-Terminal and C-Terminal Markers in Nascent Proteins  
 ; FILE REFERENCE: AMBER-03879  
 ; CURRENT APPLICATION NUMBER: US/09/382,950  
 ; CURRENT FILING DATE: 1999-08-25  
 ; NUMBER OF SEQ ID NOS: 17  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 7  
 ; LENGTH: 8  
 ; TYPE: PRT

; ORGANISM: Artificial/Unknown  
 ; FEATURE:  
 ; NAME/KEY: misc feature  
 ; LOCATION: (1)  
 ; OTHER INFORMATION: Synthetic  
 US-09-382-950-7

Query Match 81.0%; Score 34; DB 3; Length 8;  
 Best Local Similarity 71.4%; Pred. No. 4.1e+05;  
 Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 2 XHPOFER 8  
 DB 2 SHPOFER 8

RESULT 6  
 US-09-382-736B-8

; Sequence 8, Application US/09382736B  
 ; Patent No. 6306628  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Rothschild, Kenneth  
 ; APPLICANT: Gite, Sadanand  
 ; APPLICANT: Olejnik, Jerry  
 ; TITLE OF INVENTION: Methods for the Detection, Analysis and Isolation of Nascent Prot  
 ; FILE REFERENCE: AMBER-03951  
 ; CURRENT APPLICATION NUMBER: US/09/382,736B  
 ; CURRENT FILING DATE: 1999-08-25  
 ; NUMBER OF SEQ ID NOS: 18  
 ; SOFTWARE: PatentIn version 3.0  
 ; SEQ ID NO 8  
 ; LENGTH: 8  
 ; TYPE: PRT

ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Synthetic  
US-09-382-736B-8

Query Match 81.0%; Score 34; DB 3; Length 8;  
Best Local Similarity 71.4%; Pred. No. 4.1e+05;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPOPER 8  
:|||||:  
Db 2 SHPOPEK 8

RESULT 7  
US-09-619-103-9  
Sequence 9, Application US/09619103  
Patent No. 6429300  
GENERAL INFORMATION:  
APPLICANT: Kurz, Markus  
APPLICANT: Lohse, Peter  
APPLICANT: Wagner, Richard  
TITLE OF INVENTION: Peptide Acceptor Ligation Methods  
FILE REFERENCE: 50036/031002  
CURRENT APPLICATION NUMBER: US/09/619,103  
PRIOR FILING DATE: 2000-07-19  
PRIOR APPLICATION NUMBER: 60/145,834  
NUMBER OF SEQ ID NOS: 26  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 9  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: designed sequence to act as an identifying tag  
US-09-619-103-9

Query Match 81.0%; Score 34; DB 4; Length 8;  
Best Local Similarity 71.4%; Pred. No. 4.1e+05;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPOPER 8  
:|||||:  
Db 2 SHPOPEK 8

RESULT 8  
US-10-104-218-5  
Sequence 5, Application US/10104218  
Patent No. 6579705  
GENERAL INFORMATION:  
APPLICANT: Maier, Thomas  
APPLICANT: Gabbert, Carsten  
TITLE OF INVENTION: PROCESS FOR PREPARING NON-PROTEINOGENIC L-AMINO ACIDS  
FILE REFERENCE: MAIER, T. ET AL.-2  
CURRENT APPLICATION NUMBER: US/10/104,218  
PRIOR FILING DATE: 2002-03-22  
PRIOR APPLICATION NUMBER: German No. 6579705 101 16 881.0  
PRIOR FILING DATE: 2001-04-04  
PRIOR APPLICATION NUMBER: German No. 6579705 101 21 515.0  
NUMBER OF SEQ ID NOS: 5  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 5  
LENGTH: 8  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
NAME/KEY: PEPTIDE  
LOCATION: (1)..(8)  
OTHER INFORMATION: StreptagII affinity peptide for protein purification  
US-10-104-218-5

Query Match 81.0%; Score 34; DB 4; Length 8;  
Best Local Similarity 71.4%; Pred. No. 4.1e+05;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPOPER 8  
:|||||:  
Db 2 SHPOPEK 8

RESULT 9  
US-09-809-517A-9  
Sequence 9, Application US/09809517A  
Patent No. 6753136  
GENERAL INFORMATION:  
APPLICANT: Lohning, Corinna  
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b  
FILE REFERENCE: MORPHO/11  
CURRENT APPLICATION NUMBER: US/09/809,517A  
PRIOR FILING DATE: 2001-03-15  
PRIOR APPLICATION NUMBER: EP 99114072.4  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: EP 00103551.8  
NUMBER OF SEQ ID NOS: 41  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 9  
LENGTH: 8  
TYPE: PRT  
ORGANISM: artificial sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: synthetic module  
US-09-809-517A-9

Query Match 81.0%; Score 34; DB 4; Length 8;  
Best Local Similarity 71.4%; Pred. No. 4.1e+05;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPOPER 8  
:|||||:  
Db 2 SHPOPEK 8

RESULT 10  
US-09-809-517A-6  
Sequence 6, Application US/09809517A  
Patent No. 6753136  
GENERAL INFORMATION:  
APPLICANT: Lohning, Corinna  
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b  
FILE REFERENCE: MORPHO/11  
CURRENT APPLICATION NUMBER: US/09/809,517A  
PRIOR FILING DATE: 2001-03-15  
PRIOR APPLICATION NUMBER: EP 99114072.4  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: EP 00103551.8  
PRIOR FILING DATE: 2000-02-18  
NUMBER OF SEQ ID NOS: 41  
SOFTWARE: PatentIn version 3.0  
SEQ ID NO 6  
LENGTH: 10  
TYPE: PRT  
ORGANISM: artificial sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: synthetic module  
US-09-809-517A-6

Query Match 81.0%; Score 34; DB 4; Length 10;  
Best Local Similarity 71.4%; Pred. No. 1.5;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPOFER 8  
:|||||:  
Db 4 SHPOFER 10

## RESULT 11

US-08-687-559-30  
Sequence 30, Application US/08687559  
Patent No. 5955647  
GENERAL INFORMATION:  
APPLICANT: Fitch, John H.  
APPLICANT: Beachy, Roger N.  
TITLE OF INVENTION: METHOD FOR USING TOBACCO MOSAIC VIRUS TO  
TITLE OF INVENTION: OVERPRODUCE PEPTIDES AND PROTEINS  
NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Fish & Richardson, P.C.  
STREET: 4225 Executive Square, Suite 1400  
CITY: La Jolla  
STATE: CA  
COUNTRY: US  
ZIP: 92037  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Diskette  
OPERATING SYSTEM: Windows95  
SOFTWARE: FASTSEQ for Windows Version 2.0  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/687,559  
FILING DATE: No. 5955647ember 18, 1996  
CLASSIFICATION: 424  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/US95/01467  
FILING DATE: 03-FEB-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Iearn, June M.  
REGISTRATION NUMBER: 31,238  
REFERENCE/DOCKET NUMBER: 07302/011001  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 619-678-5070  
TELEFAX: 619-678-5099  
INFORMATION FOR SEQ ID NO: 30:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 13 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-687-559-30

Query Match 81.0%; Score 34; DB 2; Length 13;  
Best Local Similarity 71.4%; Pred. No. 2;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPOFER 8  
:|||||:  
Db 7 GHPOFOR 13

RESULT 12  
US-09-401-415-30  
Sequence 30, Application US/09401415  
Patent No. 6503732

## GENERAL INFORMATION:

APPLICANT: THE SCRIPPS RESEARCH INSTITUTE  
TITLE OF INVENTION: METHOD FOR USING TOBACCO MOSAIC VIRUS TO  
OVERPRODUCE PEPTIDES AND PROTEINS

NUMBER OF SEQUENCES: 30  
CORRESPONDENCE ADDRESS:

ADDRESSEE: Spensley Horn Jubas & Lubitz  
STREET: 1880 Century Park East, Suite 500  
CITY: Los Angeles  
STATE: California  
COUNTRY: USA

ZIP: 90067

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/401,415

FILING DATE: 21-Sep-1999

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/US95/01467

FILING DATE: <Unknown>

ATTORNEY/AGENT INFORMATION:

NAME: Boetlich, June M.

REGISTRATION NUMBER: 31,238

REFERENCE/DOCKET NUMBER: PD-4074

TELECOMMUNICATION INFORMATION:

TELEPHONE: (619) 455-5100

TELEFAX: (619) 455-5110

INFORMATION FOR SEQ ID NO: 30:

SEQUENCE CHARACTERISTICS:

LENGTH: 13 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

FEATURE:

NAME/KEY: Peptide

LOCATION: 1..13

SEQUENCE DESCRIPTION: SEQ ID NO: 30:

US-09-401-415-30

Query Match 81.0%; Score 34; DB 4; Length 13;  
Best Local Similarity 71.4%; Pred. No. 2;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPOFER 8  
:|||||:  
Db 7 GHPOFOR 13

## RESULT 13

US-09-809-517A-30  
Sequence 30, Application US/09809517A

Patent No. 6753136

GENERAL INFORMATION:

APPLICANT: Lohning, Corinna

TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b

FILE REFERENCE: MORPHO/11

CURRENT APPLICATION NUMBER: US/09/809,517A

PRIOR FILING DATE: 2001-03-15

PRIOR APPLICATION NUMBER: EP 99114072.4

PRIOR FILING DATE: 1999-07-20

PRIOR APPLICATION NUMBER: EP 00103551.8

PRIOR FILING DATE: 2000-02-18

NUMBER OF SEQ ID NOS: 41

SOFTWARE: Patentin version 3.0

SEQ ID NO 30

LENGTH: 21

TYPE: PRT

ORGANISM: artificial sequence

FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: synthetic module

US-09-809-517A-30

Query Match 81.0%; Score 34; DB 4; Length 21;  
Best Local Similarity 71.4%; Pred. No. 3.4;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 2 XHPOFER 8  
:|||||:



Db 15 SHPOPEK 21

## RESULT 14

US-09-809-517A-33  
; Sequence 33, Application US/09809517A  
; Patent No. 6753136  
; GENERAL INFORMATION:  
; APPLICANT: Lohning, Corinna  
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on h  
; FILE REFERENCE: MORPHO/11  
; CURRENT APPLICATION NUMBER: US/09/809,517A  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: EP 99114072.4  
; PRIOR FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 33  
; LENGTH: 22  
; TYPE: PRT  
; ORGANISM: artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module  
US-09-809-517A-33

## Query Match

Best Local Similarity 81.0%; Score 34; DB 4; Length 22;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPOPER 8

Db 16 SHPOPEK 22

## RESULT 15

US-09-809-517A-31  
; Sequence 31, Application US/09809517A  
; Patent No. 6753136  
; GENERAL INFORMATION:  
; APPLICANT: Lohning, Corinna  
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on h  
; FILE REFERENCE: MORPHO/11  
; CURRENT APPLICATION NUMBER: US/09/809,517A  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: EP 99114072.4  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: EP 00103551.8  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 31  
; LENGTH: 24  
; TYPE: PRT  
; ORGANISM: artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module  
US-09-809-517A-31

## Query Match

Best Local Similarity 81.0%; Score 34; DB 4; Length 24;  
Matches 5; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 2 XHPOPER 8

Db 18 SHPOPEK 24

Search completed: March 2, 2005, 12:25:35  
Job time : 9.10569 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 15.2846 Seconds  
(without alignments)  
125.901 Million cell updates/sec

Title: SEQ10

Perfect score: 124  
Sequence: 1 wshpofekgsgwshpofek 20

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79: \*  
1: p1r1: \*  
2: p1r2: \*  
3: p1r3: \*  
4: p1r4: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	58	46.8	493	2	B72704
2	52	41.9	103	2	C85713
3	52	41.9	103	2	A90904
4	50	40.3	346	1	S35500
5	49	39.5	726	2	T35257
6	47	37.9	103	2	E90973
7	47	37.9	103	2	A85821
8	47	37.9	433	2	H83444
9	47	37.9	472	2	AH2925
10	47	37.9	472	2	F68356
11	46.5	37.5	1589	1	RGHYC5
12	46	37.1	233	1	C48560
13	46	37.1	243	3	F90197
14	46	37.1	385	2	D41732
15	46	37.1	1885	2	JQ2183
16	45.5	36.7	256	2	JU0268
17	45.5	36.7	256	2	S37149
18	45.5	36.7	256	2	A54281
19	45	36.3	391	2	A44063
20	45	36.3	418	2	T47818
21	45	36.3	517	2	S14205
22	45	36.3	580	2	H86189
23	45	36.3	718	2	C90555
24	45	36.3	725	2	S60712
25	44.5	35.9	161	2	B42627
26	44.5	35.9	468	2	T17306
27	44.5	35.9	499	2	JC6141
28	44.5	35.9	759	2	D81657
29	44	35.5	83	2	T17752

30	44	35.5	330	2	AC2071	hypothetical prote
31	44	35.5	342	2	S37596	protein kinase MCP
32	44	35.5	365	2	AD1634	X-Pro dipeptidase
33	44	35.5	507	1	QDBE41	BGLF1 protein - hu
34	44	35.5	1253	3	T21065	hypothetical prote
35	43.5	35.1	597	1	HQEC1	hydrogenase (BC 1.
36	43.5	35.1	597	2	E85632	hydrogenase-1 larg
37	43.5	35.1	597	2	A90770	hydrogenase-1 larg
38	43	34.7	55	2	I78887	hypothetical NF-1
39	43	34.7	68	2	E85918	hypothetical prote
40	43	34.7	172	2	D82642	conserved hypothet
41	43	34.7	187	2	A45878	hypothetical prote
42	43	34.7	209	2	C89005	protein T24A6.3 (I
43	43	34.7	216	2	G75128	hypothetical prote
44	43	34.7	234	2	D96592	hypothetical prote
45	43	34.7	251	2	C81816	hypothetical prote

## ALIGNMENTS

```
RESULT 1
B72704
hypothetical protein encoded within prophage CP-9330 [imported] - Escherichia coli (strain O1
C:Species: Aeropyrum pernix
C:Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 09-Jul-2004
C:Accession: B72704
R:Kawarabayashi, Y.; Hino, Y.; Horikawa, H.; Yamazaki, S.; Hatakeyama, Y.; Jin-no, K.; Takah
awa, H.; Takamiya, M.; Maeda, S.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; K
DNA Res. 6, 83-101, 1999
A:Title: Complete genome sequence of an aerobic hyper-thermophilic Crenarchaeon, Aeropyr
A:Reference number: A72450; MUID:99310339; PMID:10382966
A:Accession: B72704
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-493 <KAM>
A:Cross-references: UNIPROT:Q9YD63; DDBJ:AP000060; NID:G5104188; PIDN:BA80034.1; PID:di
A:Experimental source: strain K1
C:Gene: APE1049

Query Match
Best Local Similarity 46.8%; Score 58; DB 2; Length 493;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;

QY 1 WSHPOFEKGGSGWSHPQFE 19
      ||| ||| ||| |||
Db 333 WPNPLFVVGCGTWEHVIE 351

RESULT 2
C85713
unknown protein encoded within prophage CP-9330 [imported] - Escherichia coli (strain O1
C:Species: Escherichia coli
C:Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: C85713
R:Riperna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glaesner, J.D.; Rose, D.J.; Mayner
iller, L.; Gotelbeck, E.J.; Davis, N.W.; Lam, A.; Dimmlanta, E.; Potamouelis, K.; Apodaca,
Nature 409, 529-533, 2001
A:Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: C85713
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <STO>
A:Cross-references: UNIPROT:O8X549; GB:AE005174; NID:G12515046; PIDN:ACG56167.1; GSPDB:
C:Experimental source: strain O157:H7, substrain EDL933
C:Gene: Z2097

Query Match
Best Local Similarity 41.9%; Score 52; DB 2; Length 103;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;
```

QY 4 POFKGGGSMWHPQEK 20  
DB 8 PVRDGYGYWTHPEYK 24

RESULT 3  
A:Accession: A90904  
hypothetical protein Eca2201 [imported] - Escherichia coli (strain O157:H7, substrain RI  
C/Species: Escherichia coli  
C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
C/Accession: A90904  
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.  
DNA Res. 8, 11-22, 2001  
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
A/Reference number: A99629; MUID:21156231; PMID:11258796  
A/Accession: A90904  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-103 <RAY>  
A/Cross-references: UNIPROT:Q8X549; GB:BA000007; PIDN:BA835624.1; PID:G13361667; GSPDB:G  
A/Experimental source: strain O157:H7, substrain RIMD 0509952  
C/Genetics:  
A/Gene: Eca2201

Query Match 41.9%; Score 52; DB 2; Length 103;  
Best Local Similarity 47.1%; Pred. No. 1.4;  
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;  
QY 4 POFKGGGSMWHPQEK 20  
DB 8 PVRDGYGYWTHPEYK 24

RESULT 4  
S35500  
heterogeneous ribonuclear particle protein homolog - Caenorhabditis elegans  
N/Alternate names: heterogeneous nuclear ribonucleoprotein homolog  
C/Species: Caenorhabditis elegans  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C/Accession: S35500; J32620  
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.  
DNA Res. 8, 11-22, 2001  
A/Title: CDNA cloning of a novel heterogeneous nuclear ribonucleoprotein gene homologue  
A/Reference number: S35500; MUID:92375684; PMID:1354852  
A/Accession: S35500  
A/Molecule type: mRNA  
A/Residues: 1-346 <TWA>  
A/Cross-references: UNIPROT:Q22037; EMBL:S43152  
R/Du, Z.; Scheet, P.; Andrews, S.  
submitted to the EMBL Data Library, December 1997  
A/Description: The sequence of C. elegans cosmid F42A6.  
A/Reference number: Z21201  
A/Accession: T32620  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Residues: 1-346 <DUZ>  
A/Molecule type: DNA  
A/Cross-references: EMBL:AF038613; PIDN:AA892051.1; GSPDB:GN00022; CESP:F42A6.7  
A/Experimental source: strain Bristol N2, clone F42A6  
C/Genetics:  
A/Gene: CESP:F42A6.7  
A/Map position: 4  
A/Introns: 9/3; 255/1  
C/Superfamily: helix-stabilizing protein; ribonucleoprotein repeat homology  
F:24-90/Domain: ribonucleoprotein repeat homology <RRM1>  
F:115-161/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match 40.3%; Score 50; DB 1; Length 346;  
Best Local Similarity 53.8%; Pred. No. 10;  
Matches 7; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 1 WSHPOFKGGGSM 13

DB 271 WGGPQQQGGGGM 283

RESULT 5  
T35257  
probable - Streptomyces coelicolor  
C/Species: Streptomyces coelicolor  
C/Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 09-Jul-2004  
C/Accession: T35257  
R/Oliver, K.; Harris, D.; Bentley, S.D.; Partridge, J.; Barrell, B.G.; Rajandream, M.A.  
submitted to the EMBL Data Library, April 1999  
A/Reference number: Z21573  
A/Accession: T35257  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-726 <OLI>  
A/Cross-references: UNIPROT:Q9X7N8; EMBL:AL049587; PIDN:CAB40682.1; GSPDB:GN00070; SCOD  
A/Experimental source: strain A3(2)  
C/Genetics:  
A/Gene: SCODB:SCSF2A.15  
C/Superfamily: enoyl-CoA hydratase/3-hydroxyacyl-CoA dehydrogenase; 3-hydroxyacyl-CoA de  
F:321-605/Domain: 3-hydroxyacyl-CoA dehydrogenase homology <HCD>

Query Match 39.5%; Score 49; DB 2; Length 726;  
Best Local Similarity 44.4%; Pred. No. 31;  
Matches 8; Conservative 3; Mismatches 7; Indels 0; Gaps 0;  
QY 1 WSHPOFKGGGSMWHPQEK 18  
DB 217 WDRPGYRIPGCTPSHPK 234

RESULT 6  
E90973  
hypothetical protein Eca2757 [imported] - Escherichia coli (strain O157:H7, substrain RI  
C/Species: Escherichia coli  
C/Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
C/Accession: E90973  
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
gasaawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.  
DNA Res. 8, 11-22, 2001  
A/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno  
A/Reference number: A99629; MUID:21156231; PMID:11258796  
A/Accession: E90973  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-103 <HAY>  
A/Cross-references: UNIPROT:Q8X4V0; GB:BA000007; PIDN:BA836180.1; PID:G13362225; GSPDB:G  
A/Experimental source: strain O157:H7, substrain RIMD 0509952  
C/Genetics:  
A/Gene: Eca2757

Query Match 37.9%; Score 47; DB 2; Length 103;  
Best Local Similarity 60.0%; Pred. No. 7.3;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;  
QY 11 GWSHPOFK 20  
DB 15 GCWTHPEYK 24

RESULT 7  
A85821  
unknown protein encoded within prophage CP-933V [imported] - Escherichia coli (strain O1  
C/Species: Escherichia coli  
C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C/Accession: A85821  
R/Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimataanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.  
A/Reference number: A85480; MUID:21074935; PMID:11206551

```
A:Accession: A95821
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <STO>
A:Cross-references: UNIPROT:Q8X4V0; GB:AE005174; NID:g12516136; PIDN:AA657029.1; GSPDB:G
A:Experimental source: strain 0157:H7, substrain EDL933
C:Genetics:
A:Gene: Z3120

Query Match      37.9%; Score 47; DB 2; Length 103;
Best Local Similarity 60.0%; Pred. No. 7.3;
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy      11 GMSHPOFEK 20
      |||:|:|:|
Db      15 GCMTHPEYK 24

RESULT 8
H83444
Probable cytochrome c PA1600 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
A:Accession: H83444
R:Stover, C.K.; Pham, X.O.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Bz
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Lardig, K.; Lim,
, Lory, S.; Olson, M.V.
Nucle 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337; PMID:10984043
A:Accession: H83444
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-433 <STO>
A:Cross-references: UNIPROT:Q913C1; GB:AE004588; GB:AE004091; NID:g9947563; PIDN:AA60498
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA1600
C:Superfamily: membrane-bound alcohol dehydrogenase, cytochrome c subunit; cytochrome c6
C:Keywords: chromoprotein; heme; iron; metalloprotein
F:58, 61/Binding site: heme (Cys) (covalent) #status predicted
F:62/Binding site: heme iron (His) (axial ligand) #status predicted
F:204, 207/Binding site: heme (Cys) (covalent) #status predicted
F:208/Binding site: heme iron (His) (axial ligand) #status predicted
F:336, 339/Binding site: heme (Cys) (covalent) #status predicted
F:340/Binding site: heme iron (His) (axial ligand) #status predicted

Query Match      37.9%; Score 47; DB 2; Length 433;
Best Local Similarity 47.1%; Pred. No. 34;
Matches 8; Conservative 2; Mismatches 7; Indels 0; Gaps 0;

Qy      4 POFKGGGSMHPOFEK 20
      |||:|:|:|
Db      90 PDRAAGIGGMSYPAPER 106

RESULT 9
AH2925
Hypothetical protein Atu3006 [imported] - Agrobacterium tumefaciens (strain C58, Dupont)
C:Species: Agrobacterium tumefaciens
C:Date: 11-Jan-2002 #sequence_revision 11-Jan-2002 #text_change 09-Jul-2004
A:Accession: AH2925
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo, L
erage, G.; Gillet, W.; Grant, C.; Genthner, D.; Kutayavin, T.; Levy, R.; Li, M.; McClell
; Karp, P.; Romero, P.; Zhang, S.
Science 294, 2317-2323, 2001
A:Authors: Yoo, H.; Teo, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kamm,
eier, E.W.
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.
A:Reference number: A85577; MUID:21608550; PMID:11743193
A:Accession: AH2925
A:Status: preliminary
A:Molecule type: DNA
```

```
A:Residues: 1-472 <KUR>
A:Cross-references: UNIPROT:Q8UBU0; GB:AE008669; PIDN:AA43822.1; PID:g17741363; GSPDB:G
A:Experimental source: strain C58 (Dupont)
C:Genetics:
A:Gene: Atu3006
A:Map position: linear chromosome

Query Match      37.9%; Score 47; DB 2; Length 472;
Best Local Similarity 63.6%; Pred. No. 37;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy      5 OFEKGGGSMH 15
      |||:|:|:|
Db      202 RFEKAGGLMTH 212

RESULT 10
F98356
Hypothetical protein AGR_L_3586 [imported] - Agrobacterium tumefaciens (strain C58, Cere
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 09-Jul-2004
A:Accession: F98356
R:Goodner, B.; Hinkle, G.; Galtung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; MUID:21608551; PMID:11743194
A:Accession: F98356
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-472 <KUR>
A:Cross-references: UNIPROT:Q8UBU0; GB:AE007870; PIDN:AAK90376.1; PID:g15160420; GSPDB:G
C:Genetics:
A:Gene: AGR_L_3586
A:Map position: linear chromosome

Query Match      37.9%; Score 47; DB 2; Length 472;
Best Local Similarity 63.6%; Pred. No. 37;
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy      5 OFEKGGGSMH 15
      |||:|:|:|
Db      202 RFEKAGGLMTH 212

RESULT 11
R6BYC5
cell division control protein CDC25 - Yeast (Saccharomyces cerevisiae)
N:Alternate names: protein L2142.6; protein YJR310c
C:Species: Saccharomyces cerevisiae
C:Date: 31-Mar-1988 #sequence_revision 31-Mar-1993 #text_change 09-Jul-2004
A:Accession: A26596; S51442; A23444; S43051; S47990
R:Broek, D.; Toda, T.; Michaeli, T.; Levin, L.; Birnmeier, C.; Zoller, M.; Powers, S.;
Cell 48, 789-799, 1987
A:Title: The S. cerevisiae CDC25 gene product regulates the Ras/adenylylate cyclase pathwa
A:Reference number: A26596; MUID:87131091; PMID:3545497
A:Accession: A26596
A:Molecule type: DNA
A:Residues: 1-1589 <BRO>
A:Cross-references: UNIPROT:P04821; EMBL:M15458; NID:g171184; PIDN:AAA34478.1; PID:g1711
R:Pauley, A.
submitted to the EMBL Data Library, November 1994
A:Description: The sequence of S. cerevisiae com1d L2142.
A:Reference number: S51437
A:Accession: S51442
A:Molecule type: DNA
A:Residues: 1-1589 <PMU>
A:Cross-references: EMBL:U17247; NID:g577216; PIDN:AA67360.1; PID:g577222; GSPDB:GN0001
R:Camonis, J.H.; Kalekine, M.; Gondre, B.; Garreau, H.; Boy-Marcotte, E.; Jacquet, M.
EMBO J. 5, 375-380, 1986
A:Title: Characterization, cloning and sequence analysis of the CDC25 gene which control
A:Reference number: A23444; MUID:86220116; PMID:3011405
A:Accession: A23444
```

A:Molecule type: DNA  
 A:Residues: 1-496, Y, 498-953, 'LSVIMNLSR', 964-1589 <CAM>  
 A:Cross-references: EMBL:X03579; NID:93483; PIDN:CAA27259.1; PID:93484  
 R:Daniel, J.H.  
 Curr. Genet. 10, 879-885, 1986  
 A:Title: The CDC25 "start" gene of *Saccharomyces cerevisiae*: sequencing of the active C-terminus  
 A:Reference number: S43051; MUID:88194639; PMID:3329037  
 A:Accession: S43051  
 A:Molecule type: DNA  
 A:Residues: 877-1589 <DAN>  
 A:Cross-references: EMBL:X03579  
 C:Genetics:  
 A:Gene: SGD: CDC25; CTN1; MIPS: YLR310C  
 A:Cross-references: SGD: S0004301; MIPS: YLR310C  
 A:Map position: 12K  
 C:Function:  
 A:Description: positive control of level of cellular cAMP at the stage at which the cell enters the budding yeast CDC25; CDC25-type guanine nucleotide exchange activator hom C/Superfamily: budding yeast CDC25; CDC25-type guanine nucleotide exchange activator hom C/Keywords: cell cycle control; transmembrane protein  
 F:55-123/Domain: BH3 homology <SH3>  
 F:1301-1542/Domain: CDC25-type guanine nucleotide exchange activator homology <SOS>  
 RESULT 12  
 C48560  
 UL56 protein - human herpesvirus 1 (strain HFEM)  
 C:Species: human herpesvirus 1  
 C>Date: 17-Feb-1994 #sequence\_revision 17-Feb-1994 #text\_change 09-Jul-2004  
 C:Accession: C48560  
 R:Rosen-Wolff, A.; Frank, S.; Raab, K.; Moyal, M.; Becker, Y.; Darai, G.  
 Virus Res. 25, 189-199, 1992  
 A:Title: Determination of the coding capacity of the BamHI DNA fragment B of apathogenic A:Reference number: A48560; MUID:93070559; PMID:1332274  
 A:Accession: C48560  
 A:Molecule type: DNA  
 A:Residues: 1-233 <ROS>  
 A:Cross-references: UNIPROT:P36297  
 A:Note: sequence extracted from NCBI backbone (NCBIN:117573, NCBI:P.117577)  
 C:Genetics:  
 A:Gene: UL56  
 C:Superfamily: herpesvirus UL56 protein  
 Query Match 37.1%; Score 46; DB 1; Length 233;  
 Best Local Similarity 43.8%; Pred. No. 24;  
 Matches 7; Conservative 1; Mismatches 8; Indels 0; Gaps 0;  
 QY 1 WSHPOFKGSGWSHP 16  
 DB 108 WQPSFRKPKGLMHP 123  
 RESULT 13  
 P90197  
 hypothetical protein SSO0519 [imported] - *Sulfolobus solfataricus*  
 C:Species: *Sulfolobus solfataricus*  
 C>Date: 24-May-2001 #sequence\_revision 24-May-2001 #text\_change 09-Jul-2004  
 C:Accession: P90197  
 R:She, Q.; Singh, R.K.; Confalonieri, F.; Zivanovic, Y.; Allard, G.; Anayez, M.J.; Chan, J.; Jongs, I.; Jeffries, A.C.; Kozera, C.J.; Medina, N.; Peng, X.; Thi-Ngoc, H.P.; Redder, R.; Jongs, R.A.; Ragani, M.A.; Sengen, C.W.; Van der Oost, J.  
 submitted to Genbank, April 2001  
 A:Description: *Sulfolobus solfataricus* complete genome.  
 A:Reference number: A99139  
 A:Accession: P90197  
 A:Status: preliminary

A:Molecule type: DNA  
 A:Residues: 1-243 <KUR>  
 A:Cross-references: UNIPROT:Q9UWS5; GB:AE006641; NID:913813677; PIDN:AAK40837.1; GSPDB:C4:Genetics:  
 A:Gene: SSO0519  
 Query Match 37.1%; Score 46; DB 2; Length 243;  
 Best Local Similarity 77.8%; Pred. No. 26;  
 Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;  
 QY 3 HPOFKGSG 11  
 DB 28 HPOFDSGGG 36  
 RESULT 14  
 D41732  
 heterogeneous nuclear RNP protein - fruit fly (*Drosophila melanogaster*)  
 C:Species: *Drosophila melanogaster*  
 C>Date: 04-Mar-1993 #sequence\_revision 02-Aug-1994 #text\_change 16-Aug-2004  
 C:Accession: D41732; G48110  
 R:Matunis, E.L.; Matunis, M.J.; Dreyfuss, G.  
 J. Cell Biol. 116, 257-269, 1992  
 A:Title: Characterization of the major hnRNP proteins from *Drosophila melanogaster*.  
 A:Reference number: A41732; MUID:92112968; PMID:1730754  
 A:Accession: D41732  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-385 <MAT>  
 A:Cross-references: UNIPROT:P48809; GB:X62639; GB:S76628; NID:911041; PIDN:CAA44505.1; P8:Kim, Y.J.; Baker, B.S.  
 Mol. Cell. Biol. 13, 174-183, 1993  
 A:Title: Isolation of RNA-type RNA-binding protein genes and the analysis of their relat A:Reference number: A48110; MUID:93109300; PMID:8417324  
 A:Accession: G48110  
 A:Status: preliminary  
 A:Molecule type: mRNA  
 A:Residues: 10-53 <KIM>  
 A:Cross-references: GB:S51720; NID:9262267; PIDN:AA824628.1; PID:9262268  
 A:Note: sequence extracted from NCBI backbone (NCBIN:121150, NCBI:P.121151)  
 C:Genetics:  
 A:Gene: FLYBase:HB98DE; FLYBase:FBP7  
 A:Cross-references: FLYBase:FBgn0001215; FLYBase:FBgn0010261  
 A:Superfamily: ribonucleoprotein repeat homology <RNP>  
 F:8-74/Domain: ribonucleoprotein repeat homology <RNP>  
 F:97-163/Domain: ribonucleoprotein repeat homology <RNP>  
 Query Match 37.1%; Score 46; DB 2; Length 385;  
 Best Local Similarity 50.0%; Pred. No. 42;  
 Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;  
 QY 1 WSHPOFKGSGWS 14  
 DB 261 WCAPSGPSGSGSMN 274  
 RESULT 15  
 JQ2183  
 hypothetical 216.5K protein - apple chlorotic leaf spot virus (isolate P-205)  
 N:Alternate names: ORF 1 protein  
 C:Species: apple chlorotic leaf spot virus  
 C>Date: 30-Sep-1993 #sequence\_revision 20-Aug-1994 #text\_change 09-Jul-2004  
 C:Accession: JQ2183  
 R:Sato, K.; Yoshikawa, N.; Takahashi, T.  
 J. Gen. Virol. 74, 1927-1931, 1993  
 A:Title: Complete nucleotide sequence of the genome of an apple isolate of apple chlorotic A:Reference number: JQ2183; MUID:93389448; PMID:8376968  
 A:Accession: JQ2183  
 A:Molecule type: mRNA  
 A:Residues: 1-1885 <SAT>  
 A:Cross-references: UNIPROT:P54891; GB:D14996; NID:9434059; PIDN:BA03641.1; PID:9453240  
 C:Superfamily: eggplant mosaic virus RNA-directed RNA polymerase

C:\Keywords: ATP

Query Match 37.1%; Score 46; DB 2; Length 1885;  
Best Local Similarity 43.8%; Pred. No. 2.3e+02;  
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

Qy 1 WSHPOFEKGGGSMSP 16  
Db 1434 WSNQPREKGPWKHP 1449

Search completed: March 2, 2005, 12:28:56  
Job time : 16.2846 secs

*This Page Blank (uspto)*





```

RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB097860; BAC44887.1; -.
DR HSSP; P30803; IAZS.
DR GO; GO:0004383; F:guanylate cyclase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro; IPR001054; G_cyclase.
DR InterPro; IPR009080; tRNAyn_1a_bind.
DR Pfam; PF00211; Guanylate_cyc; 1.
DR SMART; SM00044; CYC; 1.
DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
DR Lyase.
SQ SEQUENCE 743 AA; 83251 MW; ACFSCS3E0982813A CRC64;

Query Match 49.2%; Score 61; DB 2; Length 743;
Best Local Similarity 100.0%; Pred. No. 3.6;
Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 11 GSWSHPOFEK 20
Db 734 GSWSHPOFEK 743

```

```

RESULT 3
ID O9YD63 PRELIMINARY; PRT; 493 AA.
AC O9YD63;
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Hypothetical; protein APE1049.
GN Ordered locus names=APE1049;
OS Aeropyrum pernix.
OC Archaea; Cepharchaeota; Thermoprotei; Desulfurococcaceae;
OC Desulfurococcaceae; Aeropyrum.
OX NCBI_TaxID=56636;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=KL;
RX MEDLINE=99310339; PubMed=10382966;
RA Kawarabayashi Y., Hino Y., Horikawa H., Yamazaki S., Halkawa Y.,
RA Jin-no K., Takahashi M., Sekine M., Baba S.-I., Anai A., Koseugi H.,
RA Hosoyama A., Fukui S., Nagai Y., Nishijima K., Nakazawa H.,
RA Takamiya M., Masuda S., Funahashi T., Tanaka T., Kudoh Y.,
RA Yamazaki J., Kusbida N., Oguchi A., Aoki K.-I., Kubota K.,
RA Nakamura Y., Nomura N., Sako Y., Kikuchi H.;
RT "Complete genome sequence of an aerobic hyper-thermophilic
RT crenarchaeon, Aeropyrum pernix KL.";
RL EMBL; AP000060; BAA80034.1; -.
DR PIR; B72704; B72704.
DR GO; GO:0005381; F:iron ion transporter activity; IEA.
DR GO; GO:0006827; P:high affinity iron ion transport; IEA.
DR InterPro; IPR002491; Peripla_BP.
DR InterPro; IPR009053; Peripla_BP.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR010978; tRNA binding_arm.
DR Pfam; PF01497; Peripla_BP_2; 1.
DR Complete proteome; Hypothetical protein.
KW SEQUENCE 493 AA; 52685 MW; 9AA07ED93055826E CRC64;

```

```

Query Match 46.8%; Score 58; DB 2; Length 493;
Best Local Similarity 47.4%; Pred. No. 6.1;
Matches 9; Conservative 2; Mismatches 8; Indels 0; Gaps 0;
OY 1 WSHPEKGGSGSWHPFE 19
Db 333 WNPPLFVVGGTWEHVEIE 351

```

RESULT 4  
O8A6Y2

```

ID O8A6Y2 PRELIMINARY; PRT; 507 AA.
AC O8A6Y2;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Hypothetical protein.
GN Ordered locus names=Brl1743;
OS Bacteroides thetaiotaomicron.
OC Bacterioides; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Bacteroidaceae; Bacteroides.
OX NCBI_TaxID=818;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VPI-5482 / ATCC 29148;
RX MEDLINE=22550858; PubMed=12663928; DOI=10.1126/science.1080029;
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,
RA Chiang H.C., Hooper L.V., Gordon J.I.;
RT "A genomic view of the human-bacteroides thetaiotaomicron symbiosis.";
RL Science 299:2074-2076(2003).
DR EMBL; AE016933; AA076850.1; -.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR InterPro; IPR001087; Lipase_GDSL.
DR Pfam; PF00657; Lipase_GDSL; 1.
DR Complete proteome; Hypothetical protein.
SQ SEQUENCE 507 AA; 56254 MW; B2B890D937A1E78A CRC64;

Query Match 45.2%; Score 56; DB 2; Length 507;
Best Local Similarity 57.1%; Pred. No. 12;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;
OY 3 HPOFEKGGSGSWHP 16
Db 200 HPEFPGGNDWVHP 213

RESULT 5
ID O89NL8 PRELIMINARY; PRT; 1174 AA.
AC O89NL8;
DT 01-JUN-2003 (TREMBLrel. 24, Created)
DT 01-JUN-2003 (TREMBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE B113820 protein.
GN Ordered locus names=B113820;
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
OC Bradyrhizobiaceae; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=USDA110;
RX MEDLINE=22484998; PubMed=12597275;
RA Kaneke T., Nakamura Y., Sato S., Minamisawa K., Uchiumi T.,
RA Sasaemoto S., Watanabe A., Idegawa K., Iriuchih M., Kawashina K.,
RA Kohara M., Matsumoto M., Shimo S., Tsuruoka H., Wada T., Yamada M.,
RA Tabata S.;
RT "Complete genomic sequence of nitrogen-fixing symbiotic bacterium
RT Bradyrhizobium japonicum USDA110.";
RL DNA Res. 9:189-197(2002).
DR EMBL; AP005949; BAC49085.1; -.
DR GO; GO:0004022; F:alcohol dehydrogenase activity; IEA.
DR GO; GO:0005489; F:electron transporter activity; IEA.
DR GO; GO:0016491; F:oxidoreductase activity; IEA.
DR GO; GO:0006118; P:electron transport; IEA.
DR InterPro; IPR008274; Aldxan_dh_bind.
DR InterPro; IPR000674; Aldxan_dh_hamm.
DR InterPro; IPR003219; Cytc_ach.
DR InterPro; IPR000345; Cytc_heme_BS.
DR InterPro; IPR009056; Cytochrome_c.
DR InterPro; IPR003088; Cytc_Ct.
DR InterPro; IPR005054; RNA_rec_mot.
DR Pfam; PF01315; Ald_xan_dh_C; 1.
DR Pfam; PF02738; Ald_xan_dh_C2; 3.

```

```

DR Pfam: PF00034; Cytochrom C; 2.
DR ProDom: PD01584; CytoChrom C; UNKNOWN_3.
DR PROSITE: PS00190; CYTOCHROME_C; UNKNOWN_1.
DR PROSITE: PS00030; RRM_RNP_1; UNKNOWN_1.
DR Complete proteome.
SQ SEQUENCE 1174 AA; 125102 MW; 1365696EB14A549F CRC64;

Query Match
Best Local Similarity 52.9%; Score 55; DB 2; Length 1174;
Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 4 POFEKGGSGWSHPQFEK 20
DB 837 PDVERTGIGAMSYPAFER 853

RESULT 6
O7PP48 PRELIMINARY; PRT; 168 AA.
AC O7PP48;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DE ENSANGP00000013599 (Fragment).
GN Name=ENSANG0000001110;
OS Anopheles gambiae str. PEST.
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.
OX NCBI_TaxID=180454;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=PEST;
RA Anopheles Genome Sequencing Consortium;
RL Submitted (Apr-2003) to the EMBL/GenBank/DBJ databases.
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AAB01008960; EAA1052.2; -.
FT NON_TER
SQ SEQUENCE 168 AA; 18336 MW; 28F3E857B71210C7 CRC64;

Query Match
Best Local Similarity 43.8%; Score 54; DB 2; Length 168;
Matches 7; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

OY 1 WSHPOFEKGGSGWSHP 16
DB 95 WNHPSWMNHGSAMNHP 110

RESULT 7
O7WV12 PRELIMINARY; PRT; 955 AA.
AC O7WV12;
DT 01-MAR-2004 (TREMBlrel. 26, Created)
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Glycine cleavage system P protein.
GN Name=gcvP; OrderedLocustNames=PG1305;
OS Porphyromonas gingivalis (Bacteroides gingivalis).
OC Bacteriia; Bacteroidetes; Bacteroides (class); Bacteroidales;
OC Porphyromonadaceae; Porphyromonas.
OX NCBI_TaxID=837;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=W83;
RX MEDLINE=22829667; PubMed=12949112;
RX DOI=10.1128/JB.185.18.5591-5601.2003;
RA Nelson K.E., Fleischmann R.D., DeBoy R.T., Paulsen I.T., Fouts D.E.,
RA Eiden U.A., Daugherty S.C., Dodson R.J., Durkin A.S., Gwinn M.L.,
RA Haft D.H., Kolonay J.F., Nelson W.C., Mason T.M., Tallon L., Gray J.,
RA Grainger D., Tettelin H., Dong H., Galvin J.L., Duncan M.J.,
RA Dewhirst F.E., Fraser C.M.,

```

```

RT "Complete genome sequence of the oral pathogenic bacterium
RT Porphyromonas gingivalis strain W83."
RL J. Bacteriol. 185:5591-5601(2003).
DR EMBL; AB017176; AAO66378.1; -.
DR TIGR; PG1305; -.
DR GO; GO:0005961; C:glycine dehydrogenase complex (decarboxylat. .; IEA.
DR GO; GO:0004375; F:glycine dehydrogenase (decarboxylating) act. .; IEA.
DR GO; GO:0006544; P:glycine metabolism; IEA.
DR InterPro; IPR003437; GDC-P.
DR Pfam; PF02347; GDC-P; 1.
DR TIGRFAMs; TIGR00461; gcvP; 1.
KW Complete proteome.
SQ SEQUENCE 955 AA; 105962 MW; BF0E2BBA29183205 CRC64;

Query Match
Best Local Similarity 57.1%; Score 54; DB 2; Length 955;
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

OY 3 HPQFEKGGSGWSHP 16
DB 894 HPQYEVTDNDMSHP 907

RESULT 8
O72IW0 PRELIMINARY; PRT; 447 AA.
AC O72IW0;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Putative rRNA methyl transferase.
GN OrderedLocustNames=Trc1021;
OS Thermus thermophilus (strain HB27 / ATCC BAA-163 / DSM 7039).
OC Bacteria; Deinococcus-Thermus; Deinococci; Thermales; Thermaceae;
OC Thermus.
OX NCBI_TaxID=262724;
RN (1)
RP SEQUENCE FROM N.A.
RX PubMed=15064768;
RA Henne A., Brueggemann H., Raesch C., Wieser A., Hartsch T.,
RA Liesegang H., Johann A., Lienard T., Gohl O., Martinez-Arias R.,
RA Jacob C., Starkuviene V., Schlenczek S., Dencker S., Huber R.,
RA Klenk H.-P., Kramer W., Merkl R., Gottschalk G., Fritz H.-J.;
RT "The genome sequence of the extreme thermophile Thermus
RT thermophilus."
RL Nat. Biotechnol. 22:547-553(2004).
DR EMBL; AE017304; AAS81363.1; -.
DR GO; GO:0016740; P:transferase activity; IEA.
DR InterPro; IPR002110; ANK.
DR InterPro; IPR011023; NOP2P.
DR InterPro; IPR006174; NUBB_region.
DR PRINTS; PRO1415; ANKYRIN.
DR ProDom; PD005242; NUBB_region; 1.
DR TIGRFAMs; TIGR00446; nop2p; 1.
KW Complete proteome; Transferase.
SQ SEQUENCE 447 AA; 48704 MW; 7CA484CD337A56EA CRC64;

Query Match
Best Local Similarity 34.5%; Score 53.5; DB 2; Length 447;
Matches 10; Conservative 4; Mismatches 6; Indels 9; Gaps 1;

OY 1 WSH-----POFEKGGSGWSHPQFEK 20
DB 274 WPHRLGEGHFLARFRKGGGAWSTPLER 302

RESULT 9
O8X549 PRELIMINARY; PRT; 103 AA.
AC O8X549; Q7ADV0;
DT 01-MAR-2002 (TREMBlrel. 20, Created)
DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)

```

```

DE Hypothetical protein Z2097 (Hypothetical protein Ec2201).
GN OrderedLocNames=EC2201, Z2097;
OS Escherichia coli O157:H7.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RA MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,
RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potomousis K.,
RA Apodoca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.;
RL Nature 409:529-533 (2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=O157:H7 / RIMD 050952 / EHEC;
RA MEDLINE=21156231; PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shingawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12.";
RL DNA Res. 8:111-22 (2001).
DR EMBL; AE005346; AAG56167.1; -
DR EMBL; AP02557; BAB35624.1; -
DR PIR; A90904; A90904.
DR PIR; C85713; C85713.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 103 AA; 11980 MM; 5A42A9F91CF29E6 CRC64;

Query Match
Best Local Similarity 41.9%; Score 52; DB 2; Length 103;
Matches 8; Conservative 3; Mismatches 6; Indels 0; Gaps 0;

QY 4 PPEKGGGSMHPPEK 20
Db 8 PVRDGYGYTHPEYK 24

RESULT 10
OBENG8 PRELIMINARY; PRT; 146 AA.
AC OBENG8;
DT 01-OCT-2002 (TReMBLrel. 22, Created)
DT 01-OCT-2002 (TReMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TReMBLrel. 24, Last annotation update)
DE Molybdopterin-converting factor chain 2.
GN Name=moaB; OrderedLocNames=XAC1100;
OS Xanthomonas axonopodis (pv. citri).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=92829;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=306 / ATCC 13902 / XV 101;
RA MEDLINE=22022145; PubMed=12024217; DOI=10.1038/417459a;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farf C.S., Furian L.R.,
RA Queiroz R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
RA Alves L.M.C. do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
RA Camarotte G., Camavay F., Cardozo U., Chambeiro F., Clapina L.P.,
RA Cicerelli R.M.B., Coutinho L.L., Cursino-Santos J.R., El-Dorri H.,
RA Faria J.B., Ferreira A.J.S., Ferreira R.C.C., Ferro M.I.T.,
RA Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
RA Katuyama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
RA Locati E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
RA Martins E.C., Meidants U., Menck C.F.M., Miyaki C.Y., Moon D.H.,
RA Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,

```

```

RA Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
RA Spínola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezsa R.I.D.,
RA Terubad dos Santos M., Truffi D., Tsai S.M., White F.F.,
RA Teubad J.C., Kitajima J.P.;
RT "Comparison of the genomes of two Xanthomonas pathogens with differing
RT host specificities."
RL Nature 417:459-463 (2002).
DR EMBL; AB011739; AAM35974.1; -
DR HSP; P30749; IFN0.
DR GO; GO:0006777; P-Mo-molybdopterin cofactor biosynthesis; IEA.
DR InterPro; IPR003448; Mo_biosynth_MoAB.
DR Pfam; PF02391; MoAB; 1.
KW Complete proteome.
SQ SEQUENCE 146 AA; 16139 MM; 546976F19FA46191 CRC64;

Query Match
Best Local Similarity 41.9%; Score 52; DB 2; Length 146;
Matches 6; Conservative 4; Mismatches 9; Indels 0; Gaps 0;

QY 1 WSHPEKGGGSMHPPE 19
Db 127 WKHEHYLEGDAGWLPPELQ 145

RESULT 11
Q9LVU2 PRELIMINARY; PRT; 150 AA.
AC Q9LVU2;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TReMBLrel. 17, Last annotation update)
DE Cbl|AAP26459.1.
OS Arabidopsis thaliana (mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC MEDLINE=20181125; PubMed=10718197;
RA Sato S., Nakamura Y., Kaneko T., Katoh T., Asamizu E., Kotani H.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. X. Sequence
RT features of the regions of 3,076,755 bp covered by sixty pl and TAC
RT clones."
RL DNA Res. 7:31-63 (2000).
DR EMBL; AB018117; BAA97150.1; -
SQ SEQUENCE 150 AA; 16928 MM; F2FA83C3039E644D CRC64;

Query Match
Best Local Similarity 41.1%; Score 51; DB 2; Length 150;
Matches 7; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 WSHPEKGGGSMW 13
Db 78 WMEPSYEMGGKW 90

RESULT 12
Q7ID58 PRELIMINARY; PRT; 176 AA.
AC Q7ID58;
DT 05-JUN-2004 (TReMBLrel. 27, Created)
DT 05-JUN-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUN-2004 (TReMBLrel. 27, Last annotation update)
DE CG13011 (Fragment).
OS Drosophila yakuba (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7245;
RN [1]
RP SEQUENCE FROM N.A.

```

```

RX MEDLINE=22867302; PubMed=14525923; DOI=10.1101/gr.1311003;
RA Domazet-Losed T., Tautz D.;
RT "An evolutionary analysis of orphan genes in Drosophila.";
RL Genome Res. 13:2213-2219(2003).
DR EMBL; AF532012; AAc09911.1; -
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR008253; MARVEL.
DR Pfam; PF01284; MARVEL; 1.
FT NON_TER
SQ SEQUENCE 176 AA; 19490 MW; 33599FA9F347EB99 CRC64;

Query Match 41.1%; Score 51; DB 2; Length 176;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 WSHPOFEKGGGWSHP 16
Db 130 WLPQFLVRPFGWAVP 145

RESULT 13
Q9VXD1 PRELIMINARY; PRT; 176 AA.
ID 09VXD1;
AC 09VXD1;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE CG13011-PA.
GN ORFNames=CG13011;
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
NCBI_TaxID=7227;
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amaratunga P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Mortman J.R., Yandell M.D., Chapple M., Pfeiffer B.D.,
RA Brandon R.C., Rogers Y.H., Blazer R.G., Champe C., Rhee L.X.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abiri J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bertan B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotlier P.,
RA Burris K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Doudson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Durbin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foaier C., Gabelman A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA Glodok A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwam C.,
RA Jalali W., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Lascko P., Lei J., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Maitel B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Mervulov G., Miliushina N.V., Moberly C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Nussekm D.R., Pacleb J.M.,
RA Palazolo M., Piltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shue B.C., Siden-Klamos I., Simpson M., Skupski M.P., Smith T.,
RA Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasserman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodgett, Worley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;

```

```

RT "The genome sequence of Drosophila melanogaster.";
RL Science 287:2185-2195(2000).
RN (2)
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Kronmiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Fife E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Pacleb J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svirskas R., Taber P.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scher S.E., Myers E.W., Gibbs R.A., Rubin G.M.,
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence.";
RL Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN (3)
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Kronmiller B., Carlson J., Svirskas R.,
RA Patel S., Fife E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomics perspective.";
RL Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN (4)
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochnik S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Bertan B.P.,
RA Belencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.,
RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,
RA Stapleton M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,
RA Lewis S.E.;
RT "Annotation of the Drosophila melanogaster euchromatic genome: a
RT systematic review.";
RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).
RN (5)
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.
RN (6)
RP SEQUENCE FROM N.A.
RG FlyBase;
RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.
DR EMBL; A5003503; AAF4643.1; -
DR FlyBase; FBgn0030771; CG13011.
DR GO; GO:0016020; C:membrane; IEA.
DR InterPro; IPR008253; MARVEL.
DR Pfam; PF01284; MARVEL; 1.
SQ SEQUENCE 176 AA; 19460 MW; FDCALAAD316A1D1 CRC64;

Query Match 41.1%; Score 51; DB 2; Length 176;
Best Local Similarity 50.0%; Pred. No. 20;
Matches 8; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 1 WSHPOFEKGGGWSHP 16
Db 130 WLPQFLVRPFGWAVP 145

RESULT 14
Q654S7 PRELIMINARY; PRT; 55 AA.
ID 0654S7;
AC 0654S7;
DT 25-OCT-2004 (TrEMBLrel. 28, Created)
DT 25-OCT-2004 (TrEMBLrel. 28, Last sequence update)
DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DE Hypothetical protein B1135C02.12 (Hypothetical protein
DE OSUBA0011P19.40).
GN Name=B1135C02.12; Synonym=OSUBA0011P19.40;
OR Oryza sativa (japonica cultivar-group);
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;

```

```

OC Ehirariolidae; Oryzae; Oryza.
OX NCBI_TaxID=39947;
RN
RP SEQUENCE FROM N.A.
RX PubMed=12447438;
RA Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Katayose Y.,
  Wu J., Nilmura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,
  Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,
  Okamoto K., Ando T., Aoki H., Arita K., Hamada M., Hatada C.,
  Hishikawa S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,
  Ikono M., Itoh S., Itoh T., Itoh Y., Itoh Y., Iwabuchi A., Kamiya K.,
  Karsawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,
  Kachita K., Maehara T., Mizuno H., Miyabayashi T., Mukai Y.,
  Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,
  Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,
  Shiohara T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,
  Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshinara R., Yukawa K.,
  Zhao H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y.,
  Yano M., Jiang J., Gojobori T.;
RA "The genome sequence and structure of rice chromosome 1.";
RT Nature 420:312-316(2002).
RL Nature 420:312-316(2002).
DR EMBL; AP04358; BAD45690.1; -
DR EMBL; AP04358; BAD45690.1; -
KW Hypothetical protein.
SQ SEQUENCE 55 AA; 5354 MW; FBE1DB6BEB6D433C CRC64;

Query Match 40.3%; Score 50; DB 2; Length 55;
Best Local Similarity 50.0%; Pred. No. 7.9;
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

OY 1 WSHPOFEKGGSGSWHP 16
DB 28 WAAGRGEGGGGSAAP 43
| : | | | | | | | |
| : | | | | | | | |

RESULT 15
O8PBW7 PRELIMINARY; PRT; 145 AA.
ID O8PBW7;
DT 01-OCT-2002 (TrEMBLrel. 22, Created)
DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE MolYbdopterin-converting factor chain 2.
GN Name=moae; OrderedLocustNames=XCC0999;
OS Xanthomonas campestris (pv. campestris).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Xanthomonadales;
OC Xanthomonadaceae; Xanthomonas.
OX NCBI_TaxID=340;
RN
RP SEQUENCE FROM N.A.
RX STRAIN=ATCC 33913 / NCPPB 528;
RX MEDLINE=2022145; PubMed=12024217; DOI=10.1038/417459a;
RA da Silva A.C.R., Ferro J.A., Reinach F.C., Farah C.S., Furian L.R.,
  Queglio R.B., Monteiro-Vitorello C.B., Van Sluys M.A., Almeida N.F.,
  Alves L.M.C., do Amaral A.M., Bertolini M.C., Camargo L.E.A.,
  Camarotte G., Canavan F., Cardozo J., Chamberg F., Chaplin L.P.,
  Ciccarelli R.M.B., Coutinho L.L., Cyrino-Santos J.R., El-Dorry H.,
  Faria J.B., Ferreira A.U.S., Ferreira R.C.C., Ferro M.I.T.,
  Formighieri E.F., Franco M.C., Greggio C.C., Gruber A.,
  Katayama A.M., Kishi L.T., Leite R.P., Lemos E.G.M., Lemos M.V.F.,
  Locali E.C., Machado M.A., Madeira A.M.B.N., Martinez-Rossi N.M.,
  Martins E.C., Meidanis J., Menck C.F.M., Miyaki C.Y., Moon D.H.,
  Moreira L.M., Novo M.T.M., Okura V.K., Oliveira M.C., Oliveira V.R.,
  Pereira H.A., Rossi A., Sena J.A.D., Silva C., de Souza R.F.,
  Spindola L.A.F., Takita M.A., Tamura R.E., Teixeira E.C., Tezra R.I.D.,
  Trindade dos Santos M., Truffi D., Tsai S.M., White F.F.,
  Setubal J.C., Kitajima J.P.;
RA "Comparison of the genomes of two Xanthomonas pathogens with differing
  host specificities.";
RT Nature 417:459-463(2002).
RL Nature 417:459-463(2002).
DR EMBL; AE012196; AAM40300.1; -
DR HSSP; P30749; INVJ.
DR GO; GO:0006777; P-Mo-molybdopterin cofactor biosynthesis; IEA.

```

```

DR InterPro; IPR003448; Mb_biosynth_Moae.
DR Pfam; PF02391; Moae; 1.
KW Complete proteome.
SQ SEQUENCE 145 AA; 16129 MW; 52518278E41479F1 CRC64;

Query Match 40.3%; Score 50; DB 2; Length 145;
Best Local Similarity 35.3%; Pred. No. 23;
Matches 6; Conservative 3; Mismatches 8; Indels 0; Gaps 0;

OY 1 WSHPOFEKGGSGSWHPQ 17
DB 127 WKHEHYLEGDAGWLHP 143
| : | | | | | | | |
| : | | | | | | | |

Search completed: March 2, 2005, 12:44:22
Job time : 76.3089 secs

```



CC stronger binding than a single tag, but are displaced by a competitor.  
 CC (I) does not interfere with the function of attached proteins (II) (so it  
 CC may not be essential to remove it); facilitates detection and has easily  
 CC controllable binding properties. (I) is particularly used for purifying  
 CC FP from dilute solution in batch formats (which use simpler apparatus  
 CC than column methods and result in lower loss of FP). The present sequence  
 CC is that of a streptavidin binding peptide disclosed with the invention  
 CC  
 XX

SO Sequence 36 AA;

Query Match 85.5%; Score 106; DB 6; Length 36;  
 Best Local Similarity 55.6%; Pred. No. 3.3e-07;  
 Matches 20; Conservative 0; Mismatches 0; Indels 16; Gaps 1;

QY 1 WSHPOFEK-----GGGSMHPOFEK 20  
 |||||  
 DB 1 WSHPOFEKGGGGGGGGGGGGGGGWSHPOFEK 36

#### RESULT 2

ABP60362  
 ID ABP60362 standard; peptide; 24 AA.

XX ABP60362;

XX 28-MAR-2003 (first entry)

XX Streptavidin binding peptide SEQ ID NO 3.

XX Streptavidin; protein chip; microtitre plate; detection.

XX Synthetic.

XX Key Location/Qualifiers

XX FT Misc-difference 9.16  
 /label= unknown

XX DE10113776-A1.

XX PD 02-OCT-2002.

XX PF 21-MAR-2001; 2001DE-01013776.

XX PR 21-MAR-2001; 2001DE-01013776.

XX (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN:

XX Schmidt T;

XX WPI; 2003-031166/03.

XX New isolated peptide, useful as affinity purification tag for recombinant  
 PT protein, comprises at least two high-affinity streptavidin-binding  
 PT modules.

XX Disclosure; Page 4; 18pp; German.

XX The invention relates to an isolated peptide (I) comprising at least two  
 CC individual modules separated by 0-50 amino acids, with each containing at  
 CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain  
 CC streptavidin binding modules, are useful as affinity handles for  
 CC purification of recombinant fusion proteins (FP), also for detecting FP,  
 CC e.g. on protein chips or microtitre plates. The modules in (I) bind  
 CC strongly to streptavidin, with a co-operative effect that provides  
 CC stronger binding than a single tag, but are displaced by a competitor.  
 CC (I) does not interfere with the function of attached proteins (II) (so it  
 CC may not be essential to remove it); facilitates detection and has easily  
 CC controllable binding properties. (I) is particularly used for purifying  
 CC FP from dilute solution in batch formats (which use simpler apparatus  
 CC than column methods and result in lower loss of FP). The present sequence  
 CC is that of a streptavidin binding peptide disclosed with the invention  
 CC  
 XX

SO Sequence 24 AA;

Query Match 57.3%; Score 71; DB 6; Length 24;  
 Best Local Similarity 53.6%; Pred. No. 0.012;  
 Matches 15; Conservative 0; Mismatches 1; Indels 12; Gaps 2;

QY 1 WSHPOFEKGGGS-----MSHPOFEK 20  
 |||||  
 DB 1 WSHPOFE-----KXXXXXXXXXSHPOFEK 24

#### RESULT 3

ABP46427  
 ID ABP46427 standard; protein; 659 AA.

XX ABP46427;

XX 06-APR-2001 (first entry)

XX Bilin binding-protein associated protein #8.

XX Bilin-binding protein; mutein; BBP; digoxigenin.

XX Undefined.

XX WO200075308-A1.

XX 14-DEC-2000.

XX 08-JUN-2000; 2000WO-DE001873.

XX 08-JUN-1999; 99DE-01026068.

XX (SKER/) SKERRA A.

XX Skerra A, Schlehuber S;

XX WPI; 2001-071071/08.

XX N-PEDB; AAP25712.

XX New muteins of bilin-binding protein, useful for detecting digoxigenin  
 PT being used as label in e.g. binding assays, are very selective for  
 PT digoxigenin.

XX Example 4; Page 72-74; 80pp; German.

XX This invention describes novel polypeptides (I) that are muteins of bilin  
 CC -binding protein (BBP), which can bind digoxigenin (Dig) or its  
 CC conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and  
 CC have an amino acid (aa) substitution at at least one of the positions 28,  
 CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its  
 CC fusion proteins, are used to bind, detect, determine, immobilize or  
 CC separate Dig or its conjugates with proteins, nucleic acids,  
 CC carbohydrates, other biological or synthetic macromolecules or low  
 CC molecular weight compounds, particularly in assays where Dig is being  
 CC used as a label. Compared with Dig-specific antibodies, (I) have a  
 CC simpler structure and are easier to prepare. They have very high  
 CC specificity for Dig, relative to other steroids, and fusion partners may  
 CC be attached to either end without compromising their ability to bind  
 CC ligand  
 CC  
 XX

SO Sequence 659 AA;

Query Match 55.6%; Score 69; DB 4; Length 659;  
 Best Local Similarity 100.0%; Pred. No. 0.73;  
 Matches 11; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSHPOFEKGGG 11  
 |||||  
 DB 198 WSHPOFEKGGG 208

RESULT 4  
 ABP60369



ID	ABP60369	standard; peptide; 35 AA.
XX		
AC	ABP60369;	
XX		
DT	28-MAR-2003	(first entry)
XX		
DE	Streptavidin binding peptide SEQ ID NO 10.	
XX		
KM	Streptavidin; protein chip; microtitre plate; detection.	
XX		
OS	Synthetic.	
XX		
FH	Key	Location/Qualifiers
FT	Misc-difference 9..27	//label= unknown
FT		/note= "optionally deleted for 1-15 residues"
XX		
PN	DE10113776-A1.	
PD		
PD	02-OCT-2002.	
XX		
PE	21-MAR-2001; 2001DE-01013776.	
XX		
PR	21-MAR-2001; 2001DE-01013776.	
XX		
PA	(BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.	
XX		
PI	Schmidt T;	
XX		
DR	WPI; 2003-031166/03.	
XX		
PT	New isolated peptide, useful as affinity purification tag for recombinant	
PT	protein, comprises at least two high-affinity streptavidin-binding	
PT	modules.	
XX		
PS	Claim 7; Page 16; 18pp; German.	
XX		
CC	The invention relates to an isolated peptide (I) comprising at least two	
CC	individual modules separated by 0-50 amino acids, with each containing at	
CC	least one motif His-Pro-X where X = Gln, Asn or Met: (I), which contain	
CC	streptavidin binding modules, are useful as affinity handles for	
CC	purification of recombinant fusion proteins (FP), also for detecting FP,	
CC	e.g. on protein chips or microtitre plates. The modules in (I) bind	
CC	strongly to streptavidin, with a co-operative effect that provides	
CC	stronger binding than a single tag, but are displaced by a competitor.	
CC	(I) does not interfere with the function of attached proteins (II) (so it	
CC	may not be essential to remove it); facilitates detection and has easily	
CC	controllable binding properties. (I) is particularly used for purifying	
CC	FP from dilute solution in batch formats (which use simpler apparatus	
CC	than column methods and result in lower loss of FP). The present sequence	
CC	is that of a streptavidin binding peptide disclosed with the invention	
XX		
SO	Sequence 35 AA:	
QY	1 WSHRPEKGGGS-----WSHRPEK 20	
DB	1 WSHRPE-----KXXXXXXXXXXXXXXXXXXXXWSHRPEK 35	
	Query Match 52.8%; Score 65.5; DB 6; Length 35;	
	Best Local Similarity 38.5%; Pred. No. 0.098;	
	Matches 15; Conservative 0; Mismatches 1; Indels 23; Gaps 2	
RESULT 5		
ADP87577		
ID	ADP87577	standard; protein; 772 AA.
XX		
AC	ADP87577;	
XX		
DT	23-SEP-2004	(first entry)
XX		
DE	Human NOVX polypeptide, NOV2F.	
XX		

KW	anorectic; antidiabetic; gene therapy; vaccine; obesity; diabetes;
KV	insulin resistance; hybridization probe; chromosome mapping;
KM	tissue typing; preventive medicine; pharmacogenomics; NOVX; human.
XX	
OS	Homo sapiens.
XX	
PN	WO2004056961-A2.
XX	
PD	08-JUL-2004.
XX	
PF	27-OCT-2003; 2003WO-US034114.
XX	
PR	25-OCT-2002; 2002US-0421239P.
PR	28-OCT-2002; 2002US-0421700P.
PR	31-OCT-2002; 2002US-0432776P.
PR	13-NOV-2002; 2002US-0426197P.
PR	20-DEC-2002; 2002US-0435498P.
PR	20-DEC-2002; 2002US-0435510P.
PR	21-MAR-2003; 2003US-0456812P.
XX	
PA	(CURA-) CURAGEN CORP.
XX	
PI	Bergsh C, Catterton E, Ellerman K, Ort T, Rieger D, Chaudhuri A;
DR	MP1; 2004-500293/47.
XX	N-PDB; ADP87576.
PT	New NOVX nucleic acid molecules and polypeptides useful for preventing or
PT	treating NOVX-associated disorders, e.g. diabetes, insulin resistance or
XX	obesity, and in chromosome mapping, tissue typing or pharmacogenomics.
PS	Claim 3; SEQ ID NO 32; 570bp; English.
XX	
CC	The invention relates to a novel isolated nucleic acid molecule
CC	comprising any of the nucleotide sequences (e.g. 3634, 2127 or 2120 base
CC	pairs) fully defined in the specification; or encodes any of the amino
CC	acid sequences (e.g. 698, 702 or 709 amino acids) fully defined in the
CC	specification. The invention further comprises: an isolated polypeptide
CC	comprising any of the amino acid sequences cited above; a method for
CC	identifying compounds that modulate target polypeptide activity; an
CC	antibody that immunospecifically binds to the target polypeptide, where
CC	the target polypeptide comprises any of the above-mentioned amino acid
CC	sequences, an amino acid sequence that is at least 95% identical to the
CC	above amino acid sequences, an amino acid sequence of at least one domain
CC	of the above-mentioned amino acid sequences, or an amino acid sequence
CC	that is at least 95% identical to the domain of the above amino acid
CC	sequences; a method for identifying a potential therapeutic agent for use
CC	in treatment of a pathology, where the pathology is related to aberrant
CC	expression or aberrant physiological interactions of a target polypeptide
CC	; and a method of screening for a modulator of activity of or of latency
CC	or predeposition to a pathology associated with a target polypeptide.
CC	The modulating compounds have anorectic and antidiabetic activities. The
CC	nucleic acid sequences of the invention may be used in gene therapy to
CC	treat disorders. The proteins of the invention may be used in creating a
CC	vaccine. The composition and methods are useful for identifying compounds
CC	that modulate protein activity or for diagnosing, preventing or treating
CC	diverse pathological conditions, such as obesity, diabetes or insulin
CC	resistance. The nucleic acids are also used as hybridization probes, in
CC	chromosome mapping, tissue typing, preventive medicine, and
CC	pharmacogenomics. This sequence represents a NOVX polypeptide of the
CC	invention.
XX	
SQ	Sequence 772 AA:
QY	Query Match 51.6%; Score 64; DB 8; Length 772; Best Local Similarity 68.4%; Pred. No. 4.1; Matches 13; Conservative 0; Mismatches 4; Indels 2; Gaps 1, 4 POFKKGGS--WSHPQFK 20           754 PAFSHGFSNWSHPOFK 772
DB	

```

RESULT 6
AD20243
ID AD20243 standard; protein; 845 AA.
AC
XX
AD20243;
XX
DT 18-NOV-2004 (first entry)
XX
DE Human aggrecanase ADAMTS4 mutant E362Q protein with inserted Strep tag.
XX
KW ADAMTS4; a disintegrin-like and metalloprotease;
KW thrombospondin type 1 motif 4; repolysin; zinc metalloprotease;
KW aggrecanase; osteopontin; antiinflammatory; antirheumatic; antineumatic;
KW cytoskeletal; osteoarthritis; glioma; cancer; inflammatory joint;
KW rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;
KW human; enzyme; chromosome 1q21-q23; Strep tag; mutant; mutein.
XX
OS Homo sapiens.
OS Synthetic.
OS Unidentified.
FH Key Location/Qualifiers
FT Misc-difference 362
FT /note= "Wild-type Glu substituted by Gln"
FT /note= "Strep tag peptide 2 inserted"
FT
XX
XX WO2004011637-A2.
XX
XX 05-FEB-2004.
XX
XX 29-JUL-2003; 2003WO-US023484.
XX
XX 29-JUL-2002; 2002US-0398721P.
XX
XX (AMRP ) WYETH.
XX (CORC/) CORCORAN C J.
XX (FLAN/) FLANNERY C R.
XX (ZENG/) ZENG W.
XX (RACI/) RACIE L A.
XX (MCDO/) MCDONAGH T.
XX (FREE/) FREEMAN B A.
XX (GEOR/) GEORGIADIS K E.
XX (LAVA/) LAVALLIE E R.
XX
XX Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T,
XX Freeman BA, Georgiadis KE, Lavallie ER;
XX
XX WPI; 2004-143860/14.
XX
XX New isolated, modified ADAMTS4 (aggrecanase) protein with improved
XX stability useful for identifying inhibitors of the enzyme activity for
XX treating aggrecanase-associated conditions, including osteoarthritis.
XX
XX Claim 9; SEQ ID NO 40; 117pp; English.
XX
XX The invention relates to a novel isolated, modified ADAMTS4 (a
XX disintegrin-like and metalloprotease (repolysin type) with
XX thrombospondin type 1 motif 4) protein with improved stability compared
XX to a naturally occurring, full-length ADAMTS4 protein, where the modified
XX protein differs from the naturally-occurring, full-length ADAMTS4 protein
XX by at least one amino acid. ADAMTS proteins are a subfamily of zinc
XX metalloproteases and include aggrecanases amongst their members. The
XX protein of the invention demonstrates osteopontin, antineumatic,
XX antirheumatic, antineumatic and cytoskeletal activities and may be useful
XX for treating aggrecanase-associated conditions, including osteoarthritis,
XX glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic
XX arthritis, periodontal disease and Crohn's disease. The current sequence
XX is that of the human aggrecanase ADAMTS4 mutant E362Q protein with
XX inserted Strep tag of the invention.
XX
XX Sequence, 845 AA;

```

```

Query Match 51.2%; Score 63.5; DB 8; Length 845;
Best Local Similarity 84.6%; Pred. No. 5.3;
Matches 11; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

QY 1 WSHPOFEKGGGSW 13
||| ||| ||| ||| |||
Db 520 WSHPOFEKAGG-W 531

RESULT 7
AAW93966
ID AAW93966 standard; protein; 396 AA.
AC
XX
AAW93966;
XX
DT 02-JUL-1999 (first entry)
XX
DE Plasmid pBBP20 protein fragment.
XX
XX Lipocalin; anticalin; peptide loop; leaflet structure; antibody; therapy;
XX tissue-specific surface molecule; tumour-specific surface molecule;
XX tumour imaging; bilin binding protein; phage coat protein pIII.
XX
OS Synthetic.
FH Key Location/Qualifiers
FT Peptide 1..21
FT /label= signal_peptide
FT Protein 22..396
FT /label= fusion_peptide
FT /note= "This fusion peptide is constructed from bilin
FT binding protein, a strep-tag II region and a fragment of
FT phage coat protein pIII"
FT 22..195
FT /note= "Bilin binding protein"
FT Region 196..205
FT /note= "Strep-Tag II region"
FT Misc-difference 206
FT /note= "Encoded by TAG"
FT Protein 207..396
FT /note= "phage coat protein pIII"
XX
XX WO9916873-A1.
XX
XX 08-APR-1999.
XX
XX 25-SEP-1998; 98WO-DE002898.
XX
XX 26-SEP-1997; 97DE-01042706.
XX
XX (SKER/) SKERRA A.
XX
XX Skerra A, Beste G, Schmidt F, Stibora T;
XX
XX WPI; 1999-255093/21.
XX
XX N-ESDB; AAX24099.
XX
XX New anticalins produced by mutation of lipocalin family protein.
XX
XX Disclosure; Page 67-69; 80pp; German.
XX
XX This invention describes the preparation of anticalins from polypeptides
XX of the lipocalin family by mutation of amino acids in the region of the 4
XX peptide loops (at one end of the cylindrical 'leaflet' structure).
XX Anticalins can bind to a predetermined ligand with determinable affinity.
XX Anticalins or fusions of anticalins, are used: (a) when immobilized, for
XX separation; and (b) when labeled, for identification of ligands (or their
XX fusions or conjugates). Typically they bind to tissue- or tumor-specific
XX surface molecules and can be used for tumour imaging or directly for
XX tumour therapy. Mutations can be introduced into lipocalin polypeptides
XX more easily than into antibodies, since lipocalin polypeptides have only
XX 4 (contrast 6) sequence segments. The peptide loops can tolerate amino

```

CC acid changes without significant effect on folding. Anticalins have high  
 CC specific affinity, comparable with that for antibodies. This sequence  
 CC represents a fusion polypeptide used in the method of the invention  
 XX

SO Sequence 396 AA;

Query Match 50.0%; Score 62; DB 2; Length 396;  
 Best Local Similarity 85.7%; Pred. No. 3.8;  
 Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 WSHPOFEK--GGGS 12  
 |||||  
 DB 198 WSHPOFEKXAGGS 211

RESULT 8  
 AAB46420  
 ID AAB46420 standard; protein; 396 AA.  
 XX

AC AAB46420;

DT 06-APR-2001 (first entry)

XX Bilin binding-protein associated protein.

XX Bilin-binding protein; muten; BBP; digoxigenin.

XX Unidentified.

XX WO200075308-A1.

XX 14-DEC-2000.

XX 08-JUN-2000; 2000WO-DE001873.

XX 08-JUN-1999; 99DE-01026068.

XX (SKER/) SKERRA A.

XX Skerra A, Schlehuber S;

XX MPI: 2001-071071/08.

DR N-PSDB; AAF25695.

XX New mutenins of bilin-binding protein, useful for detecting digoxigenin  
 PT being used as label in e.g. binding assays, are very selective for  
 PT digoxigenin.

PS Example 1; Page 55-56; 80pp; German.

XX This invention describes novel polypeptides (I) that are mutenins of bilin  
 CC -binding protein (BBP), which can bind digoxigenin (Dig) or its  
 CC conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and  
 CC have an amino acid (aa) substitution at at least one of the positions 28,  
 CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its  
 CC fusion proteins, are used to bind, detect, determine, immobilize or  
 CC separate Dig or its conjugates with proteins, nucleic acids,  
 CC carboxydrates, other biological or synthetic macromolecules or low  
 CC molecular weight compounds, particularly in assays where Dig is being  
 CC used as a label. Compared with Dig-specific antibodies, (I) have a  
 CC simpler structure and are easier to prepare. They have very high  
 CC specificity for Dig, relative to other steroids, and fusion partners may  
 CC be attached to either end without compromising their ability to bind  
 CC ligand

SO Sequence 396 AA;

Query Match 50.0%; Score 62; DB 4; Length 396;  
 Best Local Similarity 85.7%; Pred. No. 3.8;  
 Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 WSHPOFEK--GGGS 12  
 |||||

DB 198 WSHPOFEKXAGGS 211

RESULT 9  
 AAB46423  
 ID AAB46423 standard; protein; 396 AA.  
 XX

AC AAB46423;

DT 06-APR-2001 (first entry)

XX Bilin binding-protein associated protein #4.

XX Bilin-binding protein; muten; BBP; digoxigenin.

XX Unidentified.

XX WO200075308-A1.

XX 14-DEC-2000.

XX 08-JUN-2000; 2000WO-DE001873.

XX 08-JUN-1999; 99DE-01026068.

XX (SKER/) SKERRA A.

XX Skerra A, Schlehuber S;

XX MPI: 2001-071071/08.

DR N-PSDB; AAF25708.

XX New mutenins of bilin-binding protein, useful for detecting digoxigenin  
 PT being used as label in e.g. binding assays, are very selective for  
 PT digoxigenin.

PS Example 2; Page 62-64; 80pp; German.

XX This invention describes novel polypeptides (I) that are mutenins of bilin  
 CC -binding protein (BBP), which can bind digoxigenin (Dig) or its  
 CC conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and  
 CC have an amino acid (aa) substitution at at least one of the positions 28,  
 CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its  
 CC fusion proteins, are used to bind, detect, determine, immobilize or  
 CC separate Dig or its conjugates with proteins, nucleic acids,  
 CC carboxydrates, other biological or synthetic macromolecules or low  
 CC molecular weight compounds, particularly in assays where Dig is being  
 CC used as a label. Compared with Dig-specific antibodies, (I) have a  
 CC simpler structure and are easier to prepare. They have very high  
 CC specificity for Dig, relative to other steroids, and fusion partners may  
 CC be attached to either end without compromising their ability to bind  
 CC ligand

SO Sequence 396 AA;

Query Match 50.0%; Score 62; DB 4; Length 396;  
 Best Local Similarity 85.7%; Pred. No. 3.8;  
 Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 WSHPOFEK--GGGS 12  
 |||||

DB 198 WSHPOFEKXAGGS 211

RESULT 10

ADA27291  
 ID ADA27291 standard; protein; 400 AA.

XX ADA27291;

XX 20-NOV-2003 (first entry)

DE Plasmid pHNGAUS fragment protein.

```

XX XX Mucin; human; neutrophil gelatinase-associated lipocalin; hNGAL; rat;
KW alpha2-microglobulin-related protein; A2m; mouse; 24p3; uterocalin.
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX FH 1..21
XX FT Peptide /note="Ompa signal sequence"
XX FT Protein 22..400
XX FT /note="Mature protein; fusion protein of hNGAL, Strep-
XX FT Tag II and a fragment of phage coat protein pIII"
XX FT Protein 22..199
XX FT /note="Mature modified hNGAL"
XX FT Peptide 200..209
XX FT /note="Strep-Tag II affinity tag"
XX FT Misc-difference 210
XX FT /note="Encoded by TAG"
XX FT Protein 211..400
XX FT /note="Amino acids 217-406 of coat protein pIII"
XX
XX WO2003029462-A1.
XX
XX 10-APR-2003.
XX
XX 27-SEP-2001; 2001WO-EP011213.
XX
XX 27-SEP-2001; 2001WO-EP011213.
XX
XX (PIER-) PIERIS PROTEOLAB AG.
XX
XX Skerra A, Sphlehuber S;
XX
XX WPI; 2003-381639/36.
XX
XX N-PSDB; ADA27285.
XX
XX Generating a mutcin of a protein for validating the protein as drug
XX target by subjecting the protein to mutagenesis at sequence positions
XX PT corresponding to sequence positions of the human neutrophil gelatinase-
XX PT associated lipocalin.
XX
XX Disclosure; Page 57-58; 68pp; English.
XX
XX The present invention relates to a method for generating mutcins of human
XX CC neutrophil gelatinase-associated lipocalin (hNGAL), rat alpha2-
XX CC microglobulin-related protein (A2m) or mouse 24p3/uterocalin, by
XX CC subjecting the protein to mutagenesis. In the method of generating a
XX CC mutcin, a nucleic acid coding for the mutcins of the protein resulting
XX CC from mutagenesis is operably fused at the 3' end with a gene coding for
XX CC the coat protein pIII of a filamentous bacteriophage of the M13-family or
XX CC for a fragment of the coat protein. The present sequence is the protein
XX CC encoded by a fragment of plasmid pHNGAL5, used to illustrate the
XX CC invention. This sequence comprises human hNGAL, a Strep-Tag II and a
XX CC fragment of phage coat protein pIII.
XX
XX Sequence 400 AA;
XX
XX Query Match 50.0%; Score 62; DB 6; Length 400;
XX Best Local Similarity 85.7%; Pred. No. 3.8;
XX Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
OY 1 WSHPOFEK--GGGS 12
DB 202 WSHPOFEKQAGGS 215

```

```

XX XX Plasmid pHNGAL3 fragment protein.
XX DE Mucin; human; neutrophil gelatinase-associated lipocalin; hNGAL; rat;
XX KW alpha2-microglobulin-related protein; A2m; mouse; 24p3; uterocalin.
XX OS Synthetic.
XX
XX Key Location/Qualifiers
XX FH 1..21
XX FT Peptide /note="Signal peptide"
XX FT Protein 22..400
XX FT /note="Mature protein; fusion protein of hNGAL, Strep-
XX FT Tag II and a fragment of phage coat protein pIII"
XX FT Protein 22..199
XX FT /note="Mature hNGAL"
XX FT Peptide 200..209
XX FT /note="Strep-Tag II affinity tag"
XX FT Misc-difference 210
XX FT /note="Encoded by TAG"
XX FT Protein 211..400
XX FT /note="Amino acids 217-406 of coat protein pIII"
XX
XX WO2003029462-A1.
XX
XX 10-APR-2003.
XX
XX 27-SEP-2001; 2001WO-EP011213.
XX
XX 27-SEP-2001; 2001WO-EP011213.
XX
XX (PIER-) PIERIS PROTEOLAB AG.
XX
XX Skerra A, Schlehuber S;
XX
XX WPI; 2003-381639/36.
XX
XX N-PSDB; ADA27286.
XX
XX Generating a mutcin of a protein for validating the protein as drug
XX target by subjecting the protein to mutagenesis at sequence positions
XX PT corresponding to sequence positions of the human neutrophil gelatinase-
XX PT associated lipocalin.
XX
XX Disclosure; Page 59-61; 68pp; English.
XX
XX The present invention relates to a method for generating mutcins of human
XX CC neutrophil gelatinase-associated lipocalin (hNGAL), rat alpha2-
XX CC microglobulin-related protein (A2m) or mouse 24p3/uterocalin, by
XX CC subjecting the protein to mutagenesis. In the method of generating a
XX CC mutcin, a nucleic acid coding for the mutcins of the protein resulting
XX CC from mutagenesis is operably fused at the 3' end with a gene coding for
XX CC the coat protein pIII of a filamentous bacteriophage of the M13-family or
XX CC for a fragment of the coat protein. The present sequence is the protein
XX CC encoded by a fragment of plasmid pHNGAL3, used to illustrate the
XX CC invention. This sequence comprises human hNGAL, a Strep-Tag II and a
XX CC fragment of phage coat protein pIII.
XX
XX Sequence 400 AA;
XX
XX Query Match 50.0%; Score 62; DB 6; Length 400;
XX Best Local Similarity 85.7%; Pred. No. 3.8;
XX Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
OY 1 WSHPOFEK--GGGS 12
DB 202 WSHPOFEKQAGGS 215

```

```

RESULT 11
ID ADA27292 standard; protein; 400 AA.
XX
XX ADA27292;
XX
XX 20-NOV-2003 (first entry)

```

```

RESULT 12
ID ADA00700 standard; protein; 400 AA.
XX
XX ADA00700;
XX
XX 20-NOV-2003 (first entry)

```

```

XX 06-NOV-2003 (first entry)
DT Modified hNGAL related fusion protein SEQ ID NO:13.
XX
DE mNuclein; human neutrophil gelatinase-associated lipocalin; hNGAL;
XX rat alpha2-microglobulin-related protein; A2m; mouse 24p3/uterocalin;
XX 24p3; mutagenesis; fusion protein; cytostatic; gene therapy;
XX tumour imaging; cancer therapy.
XX
OS Synthetic.
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT Protein /label= signal
FT /note= "modified hNGAL, Strep-tag II and phage coat
FT protein pIII fragment fusion protein"
FT Region 22..400
FT /note= "mature hNGAL"
FT Region 200..209
FT /note= "Strep-tag II affinity tag"
FT Misc-difference 210
FT /note= "amber stop codon"
FT Region 211..400
FT /note= "coat protein pIII fragment 217-406"
XX
XX WO2003029463-A2.
XX
XX 10-APR-2003.
XX
XX 18-SEP-2002; 2002WO-EP010490.
XX
XX 27-SEP-2001; 2001WO-EP011213.
XX PR 16-APR-2002; 2002WO-EP004223.
XX
XX (PIER-) PIERIS PROTEOLAB AG.
XX
XX Skerra A, Schlehuber S,
XX WPI; 2003-372000/35.
XX DR N-PSDB; ADA00731.
XX
XX Generating a mutagen of a protein, e.g. human neutrophil gelatinase-
XX associated lipocalin, rat alpha2-microglobulin-related protein or mouse
XX 24p3/uterocalin, comprises subjecting the protein to mutagenesis.
XX
XX Example 3; Page 94-95, 122pp; English.
XX
XX The present invention describes a method for generating a mutagen of a
XX protein selected from a human neutrophil gelatinase-associated lipocalin
XX (hNGAL), rat alpha2-microglobulin-related protein (A2m) and a mouse
XX 24p3/uterocalin (24p3), where the mutagen has a detectable affinity to a
XX given target, comprising subjecting the protein to mutagenesis at
XX positions 33-54, 66-83, 94-106 and/or 123-136 of hNGAL, resulting in
XX mutants of the protein. Also described: (1) a mutagen of hNGAL, A2m or
XX 24p3 having detectable binding affinity to a given target, obtained by
XX the method described above; (2) a fusion protein comprising the mutagen of
XX hNGAL, A2m or 24p3, where an enzyme, a protein or a protein domain, a
XX peptide, a signal sequence and/or an affinity tag is operably fused to
XX the amino or carboxy terminus of the mutagen; (3) a nucleic acid molecule
XX comprising a sequence encoding the mutagen of hNGAL, A2m or 24p3 or the
XX fusion protein of (2); and (4) a pharmaceutical composition comprising
XX the mutagen of hNGAL, A2m or 24p3 or the fusion protein described above,
XX and a carrier. The mutagen has cytostatic activity, and can be used in
XX gene therapy. The method is useful in generating or producing a mutagen of
XX hNGAL, A2m or 24p3 or a fusion protein. The mutagen of hNGAL, A2m or 24p3
XX or the fusion protein is useful in detecting a given target by contacting
XX the mutagen with a sample suspected of containing the given target under
XX conditions that allow complex formation between the mutagen and the given
XX target, and determining the complexed mutagen by a suitable signal. The
XX given target is a protein or protein domain, a peptide, a nucleic acid

```

```

CC molecule, an organic molecule or a metal complex and the detection is
CC carried out for validation of the protein as a pharmacological drug
CC target. The mutagen may also be used in medicine, e.g. for tumour imaging
CC or directly for cancer therapy. The present sequence represents a
CC modified hNGAL, Strep-tag II and phage coat protein pIII fragment fusion
CC protein given in the exemplification of the present invention.
XX
XX
XX Sequence 400 AA;
XX
XX Query Match 50.0%; Score 62; DB 6; Length 400;
XX Best Local Similarity 85.7%; Pred. No. 3.8;
XX Matches 12; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
XX
XX 1 WSHPOFEK--GGGS 12
XX 202 WSHPOFEKXGGGS 215
XX
XX RESULT 13
XX ID ADA00707 standard; protein; 400 AA.
XX AC ADA00707;
XX XX
XX DT 06-NOV-2003 (first entry)
XX
XX DE Modified hNGAL related fusion protein SEQ ID NO:20.
XX
XX mNuclein; human neutrophil gelatinase-associated lipocalin; hNGAL;
XX rat alpha2-microglobulin-related protein; A2m; mouse 24p3/uterocalin;
XX 24p3; mutagenesis; fusion protein; cytostatic; gene therapy;
XX tumour imaging; cancer therapy.
XX
XX Synthetic.
XX OS Homo sapiens.
XX
XX FH Key Location/Qualifiers
XX FT Peptide 1..21
XX FT Protein /label= signal
XX FT /note= "modified hNGAL, Strep-tag II and phage coat
XX protein pIII fragment fusion protein"
XX FT Region 22..400
XX FT /note= "mature hNGAL"
XX FT Region 200..209
XX FT /note= "Strep-tag II affinity tag"
XX FT Misc-difference 210
XX FT /note= "amber stop codon"
XX FT Region 211..400
XX FT /note= "coat protein pIII fragment 217-406"
XX
XX WO2003029463-A2.
XX
XX 10-APR-2003.
XX
XX 18-SEP-2002; 2002WO-EP010490.
XX
XX 27-SEP-2001; 2001WO-EP011213.
XX PR 16-APR-2002; 2002WO-EP004223.
XX
XX (PIER-) PIERIS PROTEOLAB AG.
XX
XX Skerra A, Schlehuber S,
XX WPI; 2003-372000/35.
XX DR N-PSDB; ADA00706.
XX
XX Generating a mutagen of a protein, e.g. human neutrophil gelatinase-
XX associated lipocalin, rat alpha2-microglobulin-related protein or mouse
XX 24p3/uterocalin, comprises subjecting the protein to mutagenesis.
XX
XX Disclosure; Page 104-105, 122pp; English.
XX

```



```

XX ADAMTS4: a disintegrin-like and metalloprotease;
KW thrombospondin type 1 motif 4; repolysin; zinc metalloprotease;
KW aggrecanase; osteopontin; antiinflammatory; antirheumatic;
KW cytosolic; osteoarthritis; glioma; cancer; inflammatory joint;
KW rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;
KW human; chromosome 1q21-q23; enzyme.
XX
OS Homo sapiens.
OS Synthetic.
XX
PN WO200401637-A2.
XX
PD 05-FEB-2004.
XX
PF 29-JUL-2003; 2003WO-US023484.
XX
PR 29-JUL-2002; 2002US-0398721P.
XX
PA (AMHP) WYETH.
PA (CORC) CORCORAN C J.
PA (FLAN) FLANNERY C R.
PA (ZENG) ZENG W.
PA (RACI) RACIE L A.
PA (MCDON) MCDONAGH T.
PA (FREE) FREEMAN B A.
PA (GEOR) GEORGIADIS K E.
PA (LAVA) LAVALLIE E R.
XX
PI Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T;
PI Freeman BA, Georgiadis KE, Lavallie ER;
XX
XX WPI; 2004-143860/14.
XX
PT New isolated, modified ADAMTS4 (aggrecanase) protein with improved
PT stability useful for identifying inhibitors of the enzyme activity for
PT treating aggrecanase-associated conditions, including osteoarthritis.
XX
PS Claim 9; SEQ ID NO 49; 117pp; English.
XX
CC The invention relates to a novel isolated, modified ADAMTS4 (a
CC disintegrin-like and metalloprotease (repolysin type) with
CC thrombospondin type 1 motif 4) protein with improved stability compared
CC to a naturally occurring, full-length ADAMTS4 protein, where the modified
CC protein differs from the naturally-occurring, full-length ADAMTS4 protein
CC by at least one amino acid. ADAMTS proteins are a subfamily of zinc
CC metalloproteases and include aggrecanases amongst their members. The
CC protein of the invention demonstrates osteopontin, antiinflammatory,
CC antiarthritis, antirheumatic and cytostatic activities and may be useful
CC for treating aggrecanase-associated conditions, including osteoarthritis,
CC glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic
CC arthritis, periodontal disease and Crohn's disease. The current sequence
CC is that of a human aggrecanase modified ADAMTS4 (mTS4) protein of the
CC invention.
XX
SQ Sequence 646 AA;

```

Query Match 48.4%; Score 60; DB 8; Length 646;  
 Best Local Similarity 37.5%; Pred. No. 12;  
 Matches 12; Conservative 2; Mismatches 6; Indels 12; Gaps 1;

```

OY 1 WSH-----PQFEKGGGWSHPQFEK 20
Db 615 WLHRRRAQLLEILRRPWWGRKGSAMWSHPQFEK 646

```

Search completed: March 2, 2005, 13:02:47  
 Job time : 91.0813 secs

*This Page Blank (uspto)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - protein search, using SW model

Run on: March 2, 2005, 12:29:14 ; Search time 62.1951 Seconds  
(without alignments)  
105.489 Million cell updates/sec

Title: SEQ10  
Perfect score: 124  
Sequence: 1 wbpqfkgsgswbhpqfex 20

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA.\*  
1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_NEW\_PUB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*  
20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	93	75.0	36	US-10-026-578B-11	Sequence 11, Appl
2	71	57.3	24	US-10-026-578B-3	Sequence 3, Appl
3	65	52.4	36	US-10-026-578B-10	Sequence 10, Appl
4	63.5	51.2	845	US-10-628-432-40	Sequence 40, Appl
5	60	48.4	646	US-10-628-432-49	Sequence 49, Appl
6	60	48.4	858	US-10-628-432-27	Sequence 27, Appl
7	58.5	47.2	485	US-10-628-432-47	Sequence 47, Appl
8	58.5	47.2	697	US-10-628-432-24	Sequence 24, Appl
9	58	46.8	11	US-10-354-983-29	Sequence 29, Appl
10	58	46.8	11	US-10-628-432-25	Sequence 25, Appl
11	58	46.8	661	US-10-354-983-8	Sequence 8, Appl
12	58	46.8	763	US-10-358-283-15	Sequence 15, Appl
13	57	46.0	10	US-10-147-211A-20	Sequence 20, Appl

14	56	45.2	21	9	US-09-809-517A-10	Sequence 30, Appl
15	56	45.2	22	9	US-09-809-517A-33	Sequence 33, Appl
16	56	45.2	24	9	US-09-809-517A-31	Sequence 31, Appl
17	56	45.2	25	9	US-09-809-517A-34	Sequence 34, Appl
18	56	45.2	245	17	US-10-887-228A-1	Sequence 1, Appl
19	56	45.2	246	17	US-10-887-228A-9	Sequence 9, Appl
20	56	45.2	252	17	US-10-887-228A-5	Sequence 5, Appl
21	55	44.4	633	16	US-10-628-432-53	Sequence 53, Appl
22	54	43.5	362	16	US-10-437-963-170711	Sequence 170711, A
23	53.5	43.1	338	15	US-10-425-114-44353	Sequence 44353, A
24	53.5	43.1	837	15	US-10-425-114-53249	Sequence 53249, A
25	53	42.7	180	9	US-09-811-284-249	Sequence 249, App
26	52	41.9	9	9	US-09-983-067-3	Sequence 3, Appl
27	52	41.9	10	9	US-09-809-517A-6	Sequence 6, Appl
28	52	41.9	117	10	US-09-977-137A-4	Sequence 4, Appl
29	52	41.9	117	10	US-09-977-137A-5	Sequence 5, Appl
30	52	41.9	117	10	US-09-977-137A-7	Sequence 7, Appl
31	52	41.9	117	10	US-09-977-137A-8	Sequence 8, Appl
32	52	41.9	117	10	US-09-977-137A-9	Sequence 9, Appl
33	52	41.9	117	10	US-09-977-137A-10	Sequence 10, Appl
34	52	41.9	117	10	US-09-977-137A-11	Sequence 11, Appl
35	52	41.9	117	10	US-09-977-137A-12	Sequence 12, Appl
36	52	41.9	118	10	US-09-977-137A-6	Sequence 6, Appl
37	51	41.1	8	9	US-09-809-517A-9	Sequence 9, Appl
38	51	41.1	8	9	US-09-973-145-7	Sequence 7, Appl
39	51	41.1	8	12	US-09-813-197-8	Sequence 8, Appl
40	51	41.1	8	13	US-10-104-218-5	Sequence 5, Appl
41	51	41.1	8	13	US-10-208-357-9	Sequence 9, Appl
42	51	41.1	8	14	US-10-001-934-8	Sequence 8, Appl
43	51	41.1	8	14	US-10-026-578B-2	Sequence 2, Appl
44	51	41.1	8	14	US-10-026-578B-9	Sequence 9, Appl
45	51	41.1	8	14	US-10-174-368A-7	Sequence 7, Appl

## ALIGNMENTS

RESULT 1  
US-10-026-578B-11  
; Sequence 11, Application US/10026578B  
; Publication No. US20030083474A1  
; GENERAL INFORMATION:  
; APPLICANT: IBA (GmbH)  
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta  
; FILE REFERENCE: 100810.01US1  
; CURRENT APPLICATION NUMBER: US/10/026,578B  
; CURRENT FILING DATE: 2002-11-11  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: DE 101 13 776.1  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: PCT/EP01/11846  
; PRIOR FILING DATE: 2001-10-12  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 11  
; LENGTH: 36  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic Peptide  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (3)-(28)  
; OTHER INFORMATION: Of the five (GGGS) repeats between position 9 and 28, one or mor  
; OTHER INFORMATION: repeats may be missing. However, at least one (GGGS) repeat will  
; FEATURE:  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Artificial Sequence represents peptide binding module  
US-10-026-578B-11  
Query Match 75.0%; Score 93; DB 14; Length 36;  
Best Local Similarity 52.8%; Pred. No. 2e-05;

Matches 19; Conservative 0; Mismatches 1; Indels 16; Gaps 1;

Qy 1 WSHPOFEK-----GGGWSHPQFEK 20  
|||||

Db 1 WSHPOFEKGGSGGSGGSGGSGGSGGSGHPOFEK 36

### RESULT 2

US-10-026-578B-3

/ Sequence 3, Application US/10026578B

/ Publication No. US20030083474A1

/ GENERAL INFORMATION:

/ APPLICANT: IBA (GmbH)

/ APPLICANT: Schmidt, Thomas

/ TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta

/ FILE REFERENCE: 100810.01US1

/ CURRENT FILING DATE: 2002-11-11

/ PRIOR APPLICATION NUMBER: DE 101 13 776.1

/ PRIOR FILING DATE: 2001-03-21

/ PRIOR APPLICATION NUMBER: PCT/EP01/11846

/ PRIOR FILING DATE: 2001-10-12

/ NUMBER OF SEQ ID NOS: 14

/ SOFTWARE: PatentIn version 3.1

/ SEQ ID NO 3

/ LENGTH: 24

/ TYPE: PRT

/ ORGANISM: Artificial sequence

/ FEATURE:

/ NAME/KEY: MISC\_FEATURE

/ LOCATION: (9)..(9)

/ OTHER INFORMATION: X represents a single amino acid at the position indicated

/ FEATURE:

/ NAME/KEY: MISC\_FEATURE

/ LOCATION: (10)..(10)

/ OTHER INFORMATION: X represents a single amino acid at the position indicated

/ FEATURE:

/ NAME/KEY: MISC\_FEATURE

/ LOCATION: (11)..(11)

/ OTHER INFORMATION: X represents a single amino acid at the position indicated

/ FEATURE:

/ NAME/KEY: MISC\_FEATURE

/ LOCATION: (12)..(12)

/ OTHER INFORMATION: X represents a single amino acid at the position indicated

/ FEATURE:

/ NAME/KEY: MISC\_FEATURE

/ LOCATION: (13)..(13)

/ OTHER INFORMATION: X represents a single amino acid at the position indicated

/ FEATURE:

/ NAME/KEY: MISC\_FEATURE

/ LOCATION: (14)..(14)

/ OTHER INFORMATION: X represents a single amino acid at the position indicated

/ FEATURE:

/ NAME/KEY: MISC\_FEATURE

/ LOCATION: (15)..(15)

/ OTHER INFORMATION: X represents a single amino acid at the position indicated

/ FEATURE:

/ NAME/KEY: MISC\_FEATURE

/ LOCATION: (16)..(16)

/ OTHER INFORMATION: X represents a single amino acid at the position indicated

/ FEATURE:

/ NAME/KEY: misc feature

/ OTHER INFORMATION: Artificial Sequence represents peptide binding module

US-10-026-578B-3

Query Match 57.3%; Score 71; DB 14; Length 24;

Best Local Similarity 53.6%; Pred. No. 0.012;

Matches 15; Conservative 0; Mismatches 1; Indels 12; Gaps 2;

Qy 1 WSHPOFEKGGSG-----WSHPQFEK 20  
|||||

Db 1 WSHPOFE-----KXXXXXXXXXWSHPQFEK 24

### RESULT 3

US-10-026-578B-10

/ Sequence 10, Application US/10026578B

/ Publication No. US20030083474A1

/ GENERAL INFORMATION:

/ APPLICANT: IBA (GmbH)

/ APPLICANT: Schmidt, Thomas

/ TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta

/ FILE REFERENCE: 100810.01US1

/ CURRENT FILING DATE: 2002-11-11

/ PRIOR APPLICATION NUMBER: DE 101 13 776.1

/ PRIOR FILING DATE: 2001-03-21

/ PRIOR APPLICATION NUMBER: PCT/EP01/11846

/ PRIOR FILING DATE: 2001-10-12

/ NUMBER OF SEQ ID NOS: 14

/ SOFTWARE: PatentIn version 3.1

/ SEQ ID NO 10

/ LENGTH: 36

/ TYPE: PRT

/ ORGANISM: Artificial Sequence

/ FEATURE:

/ NAME/KEY: MISC\_FEATURE

/ LOCATION: (9)..(28)

/ OTHER INFORMATION: X represents a single amino acid at each of the positions indicat

/ OTHER INFORMATION: ed; some of the amino acids may be missing. Where amino acids are

/ FEATURE: missing, the total numbers of x will be no less than 5

/ NAME/KEY: misc feature

/ OTHER INFORMATION: Artificial Sequence represents peptide binding module

US-10-026-578B-10

Query Match 52.4%; Score 65; DB 14; Length 36;

Best Local Similarity 37.5%; Pred. No. 0.11; Indels 24; Gaps 2;

Matches 15; Conservative 0; Mismatches 1; Indels 24; Gaps 2;

Qy 1 WSHPOFEKGGSG-----WSHPQFEK 20  
|||||

Db 1 WSHPOFE-----KXXXXXXXXXXXXXXXXXWSHPQFEK 36

### RESULT 4

US-10-628-432-40

/ Sequence 40, Application US/10628432

/ Publication No. US20040142863A1

/ GENERAL INFORMATION:

/ APPLICANT: Wyeth

/ TITLE OF INVENTION: Modified ADAMTS4 molecules

/ FILE REFERENCE: AM101378

/ CURRENT FILING DATE: 2003-07-29

/ NUMBER OF SEQ ID NOS: 53

/ SOFTWARE: PatentIn version 3.1

/ SEQ ID NO 40

/ LENGTH: 845

/ TYPE: PRT

/ ORGANISM: Artificial

/ FEATURE:

/ OTHER INFORMATION: ADAMTS4 ASM with insertion

US-10-628-432-40

Query Match 51.2%; Score 63.5; DB 16; Length 845;

Best Local Similarity 84.6%; Pred. No. 3.5;

Matches 11; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 1 WSHPOFEKGGSGW 13  
|||||

Db 520 WSHPOFEKAGG-W 531

```
RESULT 5
US-10-628-432-49
; Sequence 49, Application US/10628432
; Publication No. US20040142863A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Modified ADAMTS4 molecules
; FILE REFERENCE: AM101378
; CURRENT APPLICATION NUMBER: US/10/628,432
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49
; LENGTH: 646
; TYPE: PRF
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: furin-processed construct E
US-10-628-432-49

Query Match
Best Local Similarity 48.4%; Score 60; DB 16; Length 646;
Matches 12; Conservative 2; Mismatches 6; Indels 12; Gaps 1;

QY 1 WSH-----PQFKGGGWSHPQFK 20
DB 615 WLHRRALQLLEILRRPFWAGKGSAMWSHPQFK 646

RESULT 6
US-10-628-432-27
; Sequence 27, Application US/10628432
; Publication No. US20040142863A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Modified ADAMTS4 molecules
; FILE REFERENCE: AM101378
; CURRENT APPLICATION NUMBER: US/10/628,432
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 858
; TYPE: PRF
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modified ADAMTS4 molecule
US-10-628-432-27

Query Match
Best Local Similarity 48.4%; Score 60; DB 16; Length 858;
Matches 12; Conservative 2; Mismatches 6; Indels 12; Gaps 1;

QY 1 WSH-----PQFKGGGWSHPQFK 20
DB 827 WLHRRALQLLEILRRPFWAGKGSAMWSHPQFK 858

RESULT 7
US-10-628-432-47
; Sequence 47, Application US/10628432
; Publication No. US20040142863A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Modified ADAMTS4 molecules
; FILE REFERENCE: AM101378
; CURRENT APPLICATION NUMBER: US/10/628,432
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47
```

```
; LENGTH: 485
; TYPE: PRF
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: furin-processed construct C
US-10-628-432-47

Query Match
Best Local Similarity 47.2%; Score 58.5; DB 16; Length 485;
Matches 12; Conservative 3; Mismatches 4; Indels 11; Gaps 1;

QY 2 SHPQFK-----GGGWSHPQFK 20
DB 456 SKKRFDKCMVCGDGGSGGSAMWSHPQFK 485

RESULT 8
US-10-628-432-24
; Sequence 24, Application US/10628432
; Publication No. US20040142863A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Modified ADAMTS4 molecules
; FILE REFERENCE: AM101378
; CURRENT APPLICATION NUMBER: US/10/628,432
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 24
; LENGTH: 697
; TYPE: PRF
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: Truncated ADAMTS4 molecule
US-10-628-432-24

Query Match
Best Local Similarity 47.2%; Score 58.5; DB 16; Length 697;
Matches 12; Conservative 3; Mismatches 4; Indels 11; Gaps 1;

QY 2 SHPQFK-----GGGWSHPQFK 20
DB 668 SKKRFDKCMVCGDGGSGGSAMWSHPQFK 697

RESULT 9
US-10-354-983-29
; Sequence 29, Application US/10354983
; Publication No. US20040044194A1
; GENERAL INFORMATION:
; APPLICANT: AGOSTINO, MICHAEL J.
; APPLICANT: CORCORAN, CHRISTOPHER
; TITLE OF INVENTION: AGGRECANASE MOLECULES
; FILE REFERENCE: 08702.0111-00000
; CURRENT APPLICATION NUMBER: US/10/354,983
; CURRENT FILING DATE: 2003-01-31
; PRIOR APPLICATION NUMBER: 60/353,680
; PRIOR FILING DATE: 2002-01-31
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 11
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: peptide sequence
US-10-354-983-29

Query Match
Best Local Similarity 46.8%; Score 58; DB 15; Length 11;
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

QY 10 GGSWSHPQFEK 20  
| : |||||  
Db 1 GSAMSHPOFEK 11

RESULT 10  
US-10-628-432-25

/ Sequence 25, Application US/10628432  
/ Publication No. US20040142863A1  
/ GENERAL INFORMATION:

/ APPLICANT: Wyeth  
/ TITLE OF INVENTION: Modified ADAMTS4 molecules  
/ FILE REFERENCE: AM101378  
/ CURRENT APPLICATION NUMBER: US/10/628,432  
/ CURRENT FILING DATE: 2003-07-29  
/ NUMBER OF SEQ ID NOS: 53  
/ SOFTWARE: PatentIn version 3.1  
/ SEQ ID NO 25  
/ LENGTH: 11  
/ TYPE: PRT  
/ ORGANISM: Artificial  
/ FEATURE:  
/ OTHER INFORMATION: Construct C tag sequence  
US-10-628-432-25

Query Match 46.8%; Score 58; DB 16; Length 11;  
Best Local Similarity 81.8%; Pred. No. 0.31;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 10 GGSWSHPQFEK 20  
| : |||||  
Db 1 GSAMSHPOFEK 11

RESULT 11  
US-10-354-983-8

/ Sequence 8, Application US/10354983  
/ Publication No. US20040044194A1  
/ GENERAL INFORMATION:  
/ APPLICANT: AGOSTINO, MICHAEL J.  
/ TITLE OF INVENTION: AGGRECANASE MOLECULES  
/ FILE REFERENCE: 08702.0111-00000  
/ CURRENT APPLICATION NUMBER: US/10/354,983  
/ CURRENT FILING DATE: 2003-01-31  
/ PRIOR APPLICATION NUMBER: 60/353,680  
/ PRIOR FILING DATE: 2002-01-31  
/ NUMBER OF SEQ ID NOS: 32  
/ SOFTWARE: PatentIn Ver. 2.1  
/ SEQ ID NO 8  
/ LENGTH: 661  
/ TYPE: PRT  
/ ORGANISM: Homo sapiens  
US-10-354-983-8

Query Match 46.8%; Score 58; DB 15; Length 661;  
Best Local Similarity 81.8%; Pred. No. 15;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 10 GGSWSHPQFEK 20  
| : |||||  
Db 651 GSAMSHPOFEK 661

RESULT 12  
US-10-358-283-15

/ Sequence 15, Application US/10358283  
/ Publication No. US20040054149A1  
/ GENERAL INFORMATION:  
/ APPLICANT: WYETH  
/ TITLE OF INVENTION: TRUNCATED AGGRECANASE MOLECULES  
/ FILE REFERENCE: 08702-0112-00000  
/ CURRENT APPLICATION NUMBER: US/10/358,283

/ CURRENT FILING DATE: 2003-02-17  
/ PRIOR APPLICATION NUMBER: 60/354,592  
/ PRIOR FILING DATE: 2002-02-05  
/ NUMBER OF SEQ ID NOS: 33  
/ SOFTWARE: PatentIn Ver. 2.1  
/ SEQ ID NO 15  
/ LENGTH: 763  
/ TYPE: PRT  
/ ORGANISM: Homo sapiens  
US-10-358-283-15

Query Match 46.8%; Score 58; DB 15; Length 763;  
Best Local Similarity 81.8%; Pred. No. 17;  
Matches 9; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 10 GGSWSHPQFEK 20  
| : |||||  
Db 753 GSAMSHPOFEK 763

RESULT 13  
US-10-147-211A-20

/ Sequence 20, Application US/10147211A  
/ Publication No. US20030235900A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Madison, Edward  
/ TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING A TRANSMEMBRANE SERINE PROTEASE 1  
/ FILE REFERENCE: 24745-1616  
/ CURRENT APPLICATION NUMBER: US/10/147,211A  
/ CURRENT FILING DATE: 2002-05-14  
/ PRIOR APPLICATION NUMBER: 60/291,001  
/ PRIOR FILING DATE: 2001-05-14  
/ NUMBER OF SEQ ID NOS: 23  
/ SOFTWARE: FastSeq for Windows Version 4.0  
/ SEQ ID NO 20  
/ LENGTH: 10  
/ TYPE: PRT  
/ ORGANISM: streptagII sequence with a one glycine linker  
US-10-147-211A-20

Query Match 46.0%; Score 57; DB 15; Length 10;  
Best Local Similarity 90.0%; Pred. No. 0.39;  
Matches 9; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 11 GGSWSHPQFEK 20  
| : |||||  
Db 1 GSAMSHPOFEK 10

RESULT 14  
US-09-809-517A-30

/ Sequence 30, Application US/09809517A  
/ Patent No. US20020034733A1  
/ GENERAL INFORMATION:  
/ APPLICANT: Lohning, Corinna  
/ TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote  
/ FILE REFERENCE: MORPHO/11  
/ CURRENT APPLICATION NUMBER: US/09/809,517A  
/ CURRENT FILING DATE: 2001-03-15  
/ PRIOR APPLICATION NUMBER: EP 99114072.4  
/ PRIOR FILING DATE: 1999-07-20  
/ PRIOR APPLICATION NUMBER: EP 00103551.8  
/ PRIOR FILING DATE: 2000-02-18  
/ NUMBER OF SEQ ID NOS: 41  
/ SOFTWARE: PatentIn version 3.0  
/ SEQ ID NO 30  
/ LENGTH: 21  
/ TYPE: PRT  
/ ORGANISM: artificial sequence  
/ FEATURE:

OTHER INFORMATION: Description of Artificial Sequence: synthetic module  
US-09-809-517A-30

Query Match 45.2%; Score 56; DB 9; Length 21;  
Best Local Similarity 81.8%; Pred. No. 1.1;  
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 GGSMSHPQPEK 20  
| | | | |  
DB 11 GAPMSHPQPEK 21

RESULT 15

US-09-809-517A-33  
Sequence 33, Application US/09809517A  
Patent No. US20020034733A1  
GENERAL INFORMATION:  
APPLICANT: Lohning, Corinna  
TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote  
FILE REFERENCE: MORPHO/11  
CURRENT FILING DATE: 2001-03-15  
PRIOR FILING DATE: 1999-07-20  
PRIOR APPLICATION NUMBER: EP 99114072.4  
PRIOR FILING DATE: 2000-02-18  
PRIOR APPLICATION NUMBER: EP 00103551.8  
NUMBER OF SEQ ID NOS: 41  
SOFTWARE: Patentin version 3.0  
SEQ ID NO 33  
LENGTH: 22  
TYPE: PRT  
ORGANISM: artificial sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: synthetic module  
US-09-809-517A-33

Query Match 45.2%; Score 56; DB 9; Length 22;  
Best Local Similarity 81.8%; Pred. No. 1.1;  
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 GGSMSHPQPEK 20  
| | | | |  
DB 12 GAPMSHPQPEK 22

Search completed: March 2, 2005, 14:18:52  
Job time : 63.1951 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 22.7642 Seconds  
(without alignments)  
65.585 Million cell updates/sec

Title: SEQ10  
Perfect score: 124  
Sequence: 1 whpqlfkgsgswahpqlf 20

Scoring table: BLOSUM62DX  
Gapop 10.0, Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Issued Patents AA:  
1: /cgn2\_6/pcodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/pcodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/pcodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/pcodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/pcodata/1/1aa/PCUS\_COMB.pep:\*  
6: /cgn2\_6/pcodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	56	45.2	21	US-09-809-517A-30	Sequence 30, Appl
2	56	45.2	22	US-09-809-517A-31	Sequence 33, Appl
3	56	45.2	24	US-09-809-517A-31	Sequence 31, Appl
4	56	45.2	25	US-09-809-517A-34	Sequence 34, Appl
5	52	41.9	10	US-09-809-517A-6	Sequence 6, Appl
6	52	41.9	11	US-09-977-137A-4	Sequence 4, Appl
7	52	41.9	11	US-09-977-137A-5	Sequence 5, Appl
8	52	41.9	11	US-09-977-137A-7	Sequence 9, Appl
9	52	41.9	11	US-09-977-137A-8	Sequence 8, Appl
10	52	41.9	11	US-09-977-137A-9	Sequence 10, Appl
11	52	41.9	11	US-09-977-137A-10	Sequence 11, Appl
12	52	41.9	11	US-09-977-137A-11	Sequence 12, Appl
13	52	41.9	11	US-09-977-137A-12	Sequence 11, Appl
14	52	41.9	11	US-09-977-137A-6	Sequence 6, Appl
15	51	41.1	8	US-08-948-097-2	Sequence 2, Appl
16	51	41.1	8	US-09-382-950-7	Sequence 7, Appl
17	51	41.1	8	US-09-382-736B-8	Sequence 8, Appl
18	51	41.1	8	US-09-619-103-9	Sequence 9, Appl
19	51	41.1	8	US-10-104-218-5	Sequence 5, Appl
20	51	41.1	8	US-09-809-517A-9	Sequence 9, Appl
21	50	40.3	129	US-09-248-796A-22626	Sequence 22626, A
22	50	40.3	662	US-08-770-761A-2	Sequence 2, Appl
23	50	40.3	662	US-08-770-761A-5	Sequence 5, Appl
24	50	40.3	705	US-08-770-761A-7	Sequence 7, Appl
25	48	38.7	155	US-08-828-741B-11	Sequence 11, Appl
26	48	38.7	155	US-09-160-567-11	Sequence 11, Appl
27	48	38.7	155	US-09-710-299-11	Sequence 11, Appl

28	48	38.7	155	4	US-09-509-031-11	Sequence 11, Appl
29	48	38.7	495	3	US-08-828-741B-4	Sequence 4, Appl
30	48	38.7	495	3	US-09-160-567-4	Sequence 4, Appl
31	48	38.7	495	4	US-09-710-299-4	Sequence 4, Appl
32	48	38.7	495	4	US-09-509-031-4	Sequence 4, Appl
33	48	38.7	642	4	US-09-949-016-10283	Sequence 10283, A
34	48	38.7	690	3	US-08-971-089-6	Sequence 6, Appl
35	48	38.7	776	4	US-09-266-225D-10	Sequence 10, Appl
36	48	38.7	777	4	US-09-949-016-6540	Sequence 6540, Ap
37	48	38.7	1858	4	US-09-902-540-12643	Sequence 12643, A
38	47.5	38.3	604	2	US-08-468-576B-12	Sequence 12, Appl
39	47.5	38.3	604	3	US-08-468-579B-12	Sequence 12, Appl
40	47.5	38.3	604	3	US-08-468-579B-12	Sequence 12, Appl
41	47	37.9	228	3	US-09-724-864-46	Sequence 10, Appl
42	47	37.9	259	4	US-09-431-887-33	Sequence 33, Appl
43	47	37.9	434	4	US-09-252-991A-25931	Sequence 25931, A
44	46.5	37.5	1589	4	US-09-356-952-4	Sequence 4, Appl
45	46	37.1	417	3	US-09-134-000C-5002	Sequence 5002, Ap

## ALIGNMENTS

```

RESULT 1
US-09-809-517A-30
; Sequence 30, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 21
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30
Query Match          45.2% Score 56; DB 4; Length 21;
Best Local Similarity 81.8% Pred. NO. 0.11;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      10 GGSWSHPQPEK 20
      | | | | | | | |
Db       11 GAPWSHPQPEK 21

RESULT 2
US-09-809-517A-33
; Sequence 33, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41

```

```

; SOFTWARE: Patentin version 3.0
; SEQ ID NO 33
; LENGTH: 22
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33

Query Match          45.2%  Score 56; DB 4; Length 22;
Best Local Similarity 81.8%  Pred. No. 0.12;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 GGSWSHPQFEK 20
    | |||||
Db 12 GAPWSHPQFEK 22

RESULT 3
US-09-809-517A-31.
; Sequence 31, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 31
; LENGTH: 24
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-31

Query Match          45.2%  Score 56; DB 4; Length 24;
Best Local Similarity 81.8%  Pred. No. 0.13;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 GGSWSHPQFEK 20
    | |||||
Db 14 GAPWSHPQFEK 24

RESULT 4
US-09-809-517A-34.
; Sequence 34, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: artificial sequence
```

```

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34

Query Match          45.2%  Score 56; DB 4; Length 25;
Best Local Similarity 81.8%  Pred. No. 0.13;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY 10 GGSWSHPQFEK 20
    | |||||
Db 15 GAPWSHPQFEK 25

RESULT 5
US-09-809-517A-6
; Sequence 6, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 6
; LENGTH: 10
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-6

Query Match          41.9%  Score 52; DB 4; Length 10;
Best Local Similarity 88.9%  Pred. No. 0.18;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 SMSHPQFEK 20
    : |||||
Db 2  AMSHPQFEK 10

RESULT 6
US-09-977-137A-4
; Sequence 4, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summeier, Anne O.
; APPLICANT: Cagliati, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4

Query Match          41.9%  Score 52; DB 4; Length 117;
Best Local Similarity 88.9%  Pred. No. 2.7;
```



Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 SMSHPQFEK 20  
:|||||

DB 109 AMSHPQFEK 117

## RESULT 7

US-09-977-137A-5

; Sequence 5, Application US/09977137A

; Patent No. 6750042

; GENERAL INFORMATION:

; APPLICANT: Summers, Anne O.

; APPLICANT: Caguiat, Jonathan

; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and

; TITLE OF INVENTION: Methods

; FILE REFERENCE: 79-00

; CURRENT APPLICATION NUMBER: US/09/977,137A

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: US 60/240,465

; PRIOR FILING DATE: 2000-10-12

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 5

; LENGTH: 117

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: chelon

US-09-977-137A-5

Query Match 41.9%; Score 52; DB 4; Length 117;

Best Local Similarity 88.9%; Pred. No. 2.7;

Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 SMSHPQFEK 20  
:|||||

DB 109 AMSHPQFEK 117

RESULT 8  
US-09-977-137A-7

; Sequence 7, Application US/09977137A

; Patent No. 6750042

; GENERAL INFORMATION:

; APPLICANT: Summers, Anne O.

; APPLICANT: Caguiat, Jonathan

; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and

; TITLE OF INVENTION: Methods

; FILE REFERENCE: 79-00

; CURRENT APPLICATION NUMBER: US/09/977,137A

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: US 60/240,465

; PRIOR FILING DATE: 2000-10-12

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 7

; LENGTH: 117

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: chelon

US-09-977-137A-7

Query Match 41.9%; Score 52; DB 4; Length 117;

Best Local Similarity 88.9%; Pred. No. 2.7;

Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 SMSHPQFEK 20  
:|||||

DB 109 AMSHPQFEK 117

RESULT 9  
US-09-977-137A-8

; Sequence 8, Application US/09977137A

; Patent No. 6750042

; GENERAL INFORMATION:

; APPLICANT: Summers, Anne O.

; APPLICANT: Caguiat, Jonathan

; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and

; TITLE OF INVENTION: Methods

; FILE REFERENCE: 79-00

; CURRENT APPLICATION NUMBER: US/09/977,137A

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: US 60/240,465

; PRIOR FILING DATE: 2000-10-12

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 8

; LENGTH: 117

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: chelon

US-09-977-137A-8

Query Match 41.9%; Score 52; DB 4; Length 117;

Best Local Similarity 88.9%; Pred. No. 2.7;

Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 SMSHPQFEK 20  
:|||||

DB 109 AMSHPQFEK 117

RESULT 10  
US-09-977-137A-9

; Sequence 9, Application US/09977137A

; Patent No. 6750042

; GENERAL INFORMATION:

; APPLICANT: Summers, Anne O.

; APPLICANT: Caguiat, Jonathan

; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and

; TITLE OF INVENTION: Methods

; FILE REFERENCE: 79-00

; CURRENT APPLICATION NUMBER: US/09/977,137A

; CURRENT FILING DATE: 2001-10-12

; PRIOR APPLICATION NUMBER: US 60/240,465

; PRIOR FILING DATE: 2000-10-12

; NUMBER OF SEQ ID NOS: 18

; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 9

; LENGTH: 117

; TYPE: PRT

; ORGANISM: Artificial Sequence

; FEATURE:

; OTHER INFORMATION: Description of Artificial Sequence: chelon

US-09-977-137A-9

Query Match 41.9%; Score 52; DB 4; Length 117;

Best Local Similarity 88.9%; Pred. No. 2.7;

Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 SMSHPQFEK 20  
:|||||

DB 109 AMSHPQFEK 117

RESULT 11  
US-09-977-137A-10

; Sequence 10, Application US/09977137A

; Patent No. 6750042

; GENERAL INFORMATION:

; APPLICANT: Summers, Anne O.

; APPLICANT: Caguiat, Jonathan

```

; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-10

Query Match
Best Local Similarity 41.9%; Score 52; DB 4; Length 117;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 12 SWSHPQFEK 20
Db 109 AMSHPQFEK 117

RESULT 12
US-09-977-137A-11
; Sequence 11, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-11

Query Match
Best Local Similarity 41.9%; Score 52; DB 4; Length 117;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 12 SWSHPQFEK 20
Db 109 AMSHPQFEK 117

RESULT 13
US-09-977-137A-12
; Sequence 12, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliati, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
```

```

; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-12

Query Match
Best Local Similarity 41.9%; Score 52; DB 4; Length 117;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 12 SWSHPQFEK 20
Db 109 AMSHPQFEK 117

RESULT 14
US-09-977-137A-6
; Sequence 6, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliati, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-6

Query Match
Best Local Similarity 41.9%; Score 52; DB 4; Length 118;
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 12 SWSHPQFEK 20
Db 110 AMSHPQFEK 118

RESULT 15
US-08-948-097-2
; Sequence 2, Application US/08948097C
; Patent No. 6103493
; GENERAL INFORMATION:
; APPLICANT: Skerra, Arne
; APPLICANT: Voss, Selma
; TITLE OF INVENTION: Streptavidin Mutelins
; FILE REFERENCE: HUBR 119
; CURRENT APPLICATION NUMBER: US/08/948,097C
; CURRENT FILING DATE: 1997-10-09
; EARLIER APPLICATION NUMBER: DE 196 41 876.3
; EARLIER FILING DATE: 1996-10-10
; NUMBER OF SEQ ID NOS: 17
; SEQ ID NO 2
; LENGTH: 8
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; NAME/KEY: BINDING
; OTHER INFORMATION: Binding ligand for streptavidin
```

US-08-948-097-2

Query Match 41.1%; Score 51; DB 3; Length 8;  
Best Local Similarity 100.0%; Pred. No. 4.1e+05;  
Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 1 WSHPOFEK 8  
|||  
Db 1 WSHPOFEK 8

Search completed: March 2, 2005, 12:25:36  
Job time : 23.7642 secs

***This Page Blank (uspto)***

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 18.3415 Seconds  
(without alignments)  
125.901 Million cell updates/sec

Title: SEQ11

Perfect score: 146

Sequence: 1 webpofekgsgsgsgswahpafek 24

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : PIR 79:\*

1: PIR1:\*

2: PIR2:\*

3: PIR3:\*

4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	41.8	99	2	AB2153 RNA binding protei
2	60	41.1	448	1	A56018 transcription fact
3	58	39.7	449	1	S30205 transcription fact
4	58	39.7	451	1	A40168 transcription fact
5	58	39.7	690	2	D98318 ABC protein AGR 1
6	58	39.7	1079	2	F84946 cardamoyl-phosphat
7	57.5	39.4	161	2	B42627 cement precursor p
8	56.5	38.7	346	1	S35500 heterogeneous ribo
9	56	38.4	621	2	T20863 hypothetical prote
10	55	37.7	589	2	S71954 RNA/spDNA-binding
11	54	37.0	333	2	E48423 homeotic protein e
12	54	37.0	447	2	F81678 GTP-binding protei
13	54	37.0	447	2	G71521 probable GTP bindi
14	54	37.0	732	2	T08420 1-phosphatidylinos
15	52	35.6	154	2	E96603 unknown protein F1
16	52	35.6	155	2	T35626 probable membrane
17	52	35.6	425	2	F70038 arabinogalactan en
18	52	35.6	976	1	TWMSMD macrophage colony-
19	52	35.6	978	2	S16385 macrophage colony-
20	51	34.9	158	2	F96692 glycolytic-rich prote
21	51	34.9	169	1	S38331 glycine-rich RNA-b
22	51	34.9	311	2	T02783 probable homeotic
23	51	34.9	337	2	G84590 probable heat shoc
24	51	34.9	358	2	A54265 CCAAT/enhancer-bin
25	51	34.9	365	2	A26459 helix-destabilizin
26	51	34.9	394	2	C64112 heat shock protein
27	51	34.9	466	2	AH1126 endo-1,4-beta-xyla
28	51	34.9	592	2	E82759 endo-1,4-beta-gluc
29	51	34.9	1240	2	JC5209 insulin receptor 8

30	51	34.9	1407	2	B42239 adenylyate cyclase
31	50.5	34.6	256	2	JU0268 major prion protei
32	50.5	34.6	256	2	S37149 prion protein - go
33	50.5	34.6	256	2	A54281 major prion protei
34	50.5	34.6	364	1	QOBE9 BORE1 protein - hu
35	50	34.2	236	1	XMBST9 transcription acti
36	50	34.2	243	2	F90197 hypothetical prote
37	50	34.2	252	2	S53631 major prion protei
38	50	34.2	254	2	A23544 major prion protei
39	50	34.2	288	2	T21222 hypothetical prote
40	50	34.2	282	2	S68798 RNA-binding protei
41	50	34.2	311	2	F86341 hypothetical prote
42	50	34.2	434	2	T01013 hypothetical prote
43	50	34.2	545	2	A38447 oligopeptide ABC t
44	50	34.2	1047	2	G87398 AcRb/AcrF fam
45	49.5	33.9	165	2	T24470 hypothetical prote

#### ALIGNMENTS

RESULT 1  
AB2153  
RNA binding protein [imported] - Noctoc sp. (strain PCC 7120)  
C:Species: Noctoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
A:Note: Noctoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #text\_change 09-Jul-2004  
C:Accession: AB2153  
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Kanehisa, K.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yasuda, M.; Tabata, S.  
DNA Res. 6, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120  
A:Reference number: AB1807; MUID:21595285; PMID:11759840  
A:Accession: AB2153  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-99 <RUR>  
A:Cross-references: UNIPROT:Q9WX37; GB:BA000019; PIDN:BAW74476.1; PID:gl7131870; GSPDB:1A56018  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: rbpE  
C:Superfamily: unassigned ribonucleoprotein repeat-containing proteins; ribonucleoprotein repeat-containing proteins  
Query Match 41.8%; Score 61; DB 2; Length 99;  
Best Local Similarity 64.7%; Pred. No. 0.83;  
Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;  
QY 2 SHPOFEKGGSGSGSWS 18  
DB 77 ARPREKGRSGSGSWS 93  
RESULT 2  
A56018  
transcription factor Oct-6 - human  
W:Alternate names: transcription factor SCIP  
C:Species: Homo sapiens (man)  
C:Date: 05-May-1995 #sequence\_revision 05-May-1995 #text\_change 09-Jul-2004  
C:Accession: A56018; S30218  
R:Paauw, I.; Hsu, H.J.; Fuchs, E.  
Mol. Cell. Biol. 14, 3263-3275, 1994  
A:Title: Oct-6: a regulator of keratinocyte gene expression in stratified squamous epithelium  
A:Reference number: A56018; MUID:94217723; PMID:7909356  
A:Accession: A56018  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-448 <FAU>  
A:Cross-references: UNIPROT:Q03052; GB:L26494; MTD:G508989; PIDN:AAA59965.1; PID:G508990  
A:Note: It is uncertain whether Met-1 or Met-51 is the initiator  
R:Tobler, A.; Schreiber, B.; Fontana, A.  
Nucleic Acids Res. 21, 1043, 1993  
A:Title: The human Oct-6 POU transcription factor lacks the first 50 amino acids of its  
A:Reference number: S30218; MUID:93197134; PMID:8451175

A/Accession: S30218  
 A/Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A/Molecule type: mRNA  
 A/Residues: 51-448 <TOB>  
 A/Cross-references: EMBL:118284; NID:935133; PTDN:CAA79158.1; PID:935134  
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1992  
 C/Genetics:  
 A/Gene: GDB:POU3F1; OCT-6; OTF6  
 A/Cross-references: GDB:138779  
 A/Map position: 19pter-19qter  
 C/Superfamily: transcription factor Brn-1; homeobox homology; POU domain homology  
 C/Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 F/251-318/Domain: POU domain homology <POU>  
 F/337-393/Domain: homeobox homology <HOX>

Query Match 41.1%; Score 60; DB 1; Length 448;  
 Best Local Similarity 45.5%; Pred. No. 4.8;  
 Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Qy 2 SHPQFKGGGGGSGGWSHPQFE 23  
 Db 67 AHPQWLPTGGGGGDMAGHALE 88

RESULT 3  
 S30205  
 transcription factor Oct-6 - mouse  
 N/Alternate names: class III POU domain protein SCIP  
 C/Species: Mus musculus (house mouse)  
 C/Date: 02-Dec-1993 #sequence revision 26-May-1995 #text\_change 09-Jul-2004  
 A/Accession: S30205; S31226; S13083; S11999  
 R/Zimmerman, E.C.; Jones, C.M.; Pet, V.; Hogan, B.L.M.; Magnuson, M.A.  
 Nucleic Acid Res. 19, 956, 1991  
 A/Title: Nucleotide sequence of mouse SCIP cDNA, a POU-domain transcription factor.  
 A/Reference number: S30205; MUID:91204458; PMID:1840678  
 A/Accession: S30205  
 A/Status: nucleic acid sequence not shown; translation not shown  
 A/Molecule type: mRNA  
 A/Residues: 1-449 <ZIM>  
 A/Cross-references: UNIPROT:P21952; EMBL:X56959; NID:943681; PTDN:CAA40280.1; PID:943682  
 A/Note: the nucleotide sequence was submitted to the EMBL Data Library, December 1990  
 R/Hara, Y.; Rovescalli, A.C.; Kim, Y.; Nirenberg, M.  
 Proc. Natl. Acad. Sci. U.S.A. 89, 3280-3284, 1992  
 A/Title: Structure and evolution of four POU domain genes expressed in mouse brain.  
 A/Reference number: S31223; MUID:92228768; PMID:1565620  
 A/Accession: S31226  
 A/Molecule type: DNA  
 A/Residues: 1-449 <HAR>  
 A/Cross-references: EMBL:W8302; NID:9200450; PTDN:AAA39963.1; PID:9200451  
 R/Meijer, D.; Graus, A.; Kraay, R.; Langeveld, A.; Mulder, M.P.; Grosveld, G.  
 Nucleic Acid Res. 18, 7357-7365, 1990  
 A/Title: The octamer binding factor Oct6: cDNA cloning and expression in early embryonic  
 A/Reference number: S13083; MUID:91081324; PMID:1979677  
 A/Accession: S13083  
 A/Molecule type: mRNA  
 A/Residues: 1-449 <MEI>  
 A/Cross-references: EMBL:X54628; NID:953505; PTDN:CAA38445.1; PID:953506  
 R/Suzuki, N.; Koidenohda, H.; Neuman, T.; Grunse, F.; Schneider, H.R.  
 EMBO J. 9, 3723-3732, 1990  
 A/Title: Oct-6: a POU transcription factor expressed in embryonal stem cells and in the  
 A/Reference number: S11999; MUID:91006074; PMID:1976514  
 A/Accession: S11999  
 A/Molecule type: mRNA  
 A/Residues: 1-26; 28-449 <SUZ>  
 A/Cross-references: EMBL:X57482; NID:953507; PTDN:CAA40720.1; PID:953508  
 C/Superfamily: transcription factor Brn-1; homeobox homology; POU domain homology  
 C/Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 F/252-319/Domain: POU domain homology <POU>  
 F/338-394/Domain: homeobox homology <HOX>

Query Match 39.7%; Score 58; DB 1; Length 449;  
 Best Local Similarity 52.9%; Pred. No. 8.5;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SHPQFKGGGGGSGGWS 18  
 Db 67 AHPQWLPTGGGGGDMWA 83

RESULT 4  
 A40168  
 transcription factor Oct-6 - rat  
 N/Alternate names: POU domain protein Tct-1; transcription factor SCIP  
 C/Species: Rattus norvegicus (Norway rat)  
 C/Date: 05-Jun-1992 #sequence revision 05-Jun-1992 #text\_change 09-Jul-2004  
 A/Accession: A40168; A39694; S05449; S05044  
 R/Monuki, E.S.; Kohn, R.; Weinmaster, G.; Trapp, B.D.; Lenke, G.  
 Science 249, 1300-1303, 1990  
 A/Title: Expression and activity of the POU transcription factor SCIP.  
 A/Reference number: A40168; MUID:90378306; PMID:1975954  
 A/Accession: A40168  
 A/Status: preliminary  
 A/Molecule type: mRNA  
 A/Residues: 1-451 <MON>  
 A/Cross-references: UNIPROT:P20267; GB:M72711; GB:M35205; NID:9206866; PTDN:AAA42118.1;  
 R/He, X.; Gerreiro, R.; Simmons, D.M.; Park, R.E.; Lin, C.R.; Swanson, L.W.; Rosenfeld, M.  
 Mol. Cell. Biol. 11, 1739-1744, 1991  
 A/Title: Tct-1, a member of the POU domain gene family, binds the promoter of the gene e  
 A/Reference number: A39694; MUID:91141528; PMID:1705013  
 A/Accession: A39694  
 A/Molecule type: mRNA  
 A/Residues: 35-451 <HEA>  
 A/Cross-references: GB:M63712; NID:9207539; PTDN:AAA42303.1; PID:9207540  
 A/Note: the authors translated the codon GCT for residue 249 as pro  
 R/He, X.; Treacy, M.N.; Simmons, D.M.; Ingraham, H.A.; Swanson, L.W.; Rosenfeld, M.G.  
 Nature 340, 662, 1989  
 A/Title: Correction. Expression of a large family of POU-domain regulatory genes in mamm  
 A/Reference number: S05449  
 A/Accession: S05449  
 A/Status: preliminary; not compared with conceptual translation  
 A/Molecule type: mRNA  
 A/Residues: 265-389 <HEX>  
 A/Note: this is a revision to the sequence from reference S05042  
 R/He, X.; Treacy, M.N.; Simmons, D.M.; Ingraham, H.A.; Swanson, L.W.; Rosenfeld, M.G.  
 Nature 340, 35-42, 1989  
 A/Title: Expression of a large family of POU-domain regulatory genes in mammalian brain  
 A/Reference number: S05042; MUID:89295573; PMID:2739723  
 A/Accession: S05044  
 A/Status: not compared with conceptual translation  
 A/Molecule type: mRNA  
 A/Residues: 265-319; 'A', 321-335; 336-389 <HEF>  
 A/Note: this sequence has been revised in reference S05449  
 C/Genetics:  
 C/Superfamily: transcription factor Brn-1; homeobox homology; POU domain homology  
 C/Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 F/254-321/Domain: POU domain homology <POU>  
 F/340-396/Domain: homeobox homology <HOX>

Query Match 39.7%; Score 58; DB 1; Length 451;  
 Best Local Similarity 52.9%; Pred. No. 8.5;  
 Matches 9; Conservative 3; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SHPQFKGGGGGSGGWS 18  
 Db 69 AHPQWLPTGGGGGDMWA 85

RESULT 5  
 D98318  
 ABC protein AGR 1, 3000 (AB015053) (imported) - Agrobacterium tumefaciens (strain C58, Ce  
 C/Species: Agrobacterium tumefaciens  
 C/Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 09-Jul-2004  
 A/Accession: D98318  
 R/Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldman,  
 A.; Liu, F.; Wollam, C.; Allinger, M.; Doughly, D.; Scott, C.; Lappas, C.; Markelz, B.

Science 294, 2323-2328, 2001  
A>Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent *Agrobacterium tumefaciens*  
A'Accession: D98318  
A'Accession: D98318  
A'Molecule type: DNA  
A'Results: 1-690 <KUR>  
A'Cross-references: UNIPROT:Q8U4Q1; GB:AE007870; PDB:AAK90070.1; PDB:G15160053; GSPDB:G15160053  
A'Gene: AGR\_L3000  
A'Map position: linear chromosome

Query Match  
Best Local Similarity 39.7%; Score 58; DB 2; Length 690;  
Matches 10; Conservative 2; Mismatches 6; Indels 0; Gaps 0;

QY 7 EKGGGGGGGWSHPOFEK 24  
DB 101 DRGGGGGGGGSHSGFHK 118

RESULT 6  
F84946  
carbamoyl-phosphate synthase (glutamine-hydrolyzing) (EC 6.3.5.5) large chain [imported]  
C'Species: Buchnera sp.  
C'Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 23-Mar-2001  
C'Accession: F84946  
R'Shigenobu, S.; Matsumoto, H.; Hattori, M.; Sakaki, Y.; Ishikawa, H.  
Nucleotide 407, 81-86, 2000  
A>Title: Genome sequence of the endocellular bacterial symbiont of aphids *Buchnera* sp. A  
A'Accession: F84946  
A'Accession: F84946  
A'Status: preliminary  
A'Molecule type: DNA  
A'Results: 1-1079 <STO>  
A'Cross-references: GB:AP000398; GSPDB:GN00144  
A'Experimental source: strain APS  
C'Genetics:  
A'Gene: carb, BU144  
C'Superfamily: carbamoyl-phosphate synthase (glutamine-hydrolyzing) large chain, biotin  
C'Keywords: ligase

Query Match  
Best Local Similarity 39.7%; Score 58; DB 2; Length 1079;  
Matches 10; Conservative 4; Mismatches 7; Indels 0; Gaps 0;

QY 4 POFKGGGGGGWSHPOFEK 24  
DB 170 PSFTGGGGGGTAVNHBEFEK 190

RESULT 7  
B42627  
cement precursor protein PC-2 - polychaete (*Phragmatopoma californica*) (fragments)  
C'Species: *Phragmatopoma californica*  
C'Date: 03-Feb-1994 #sequence\_revision 03-Feb-1994 #text\_change 03-Jul-2004  
C'Accession: B42627  
R'Walte, J.H.; Jensen, R.A.; Morse, D.E.  
Biochemistry 31, 5733-5738, 1992  
A>Title: Cement precursor proteins of the reef-building polychaete *Phragmatopoma californica*  
A'Accession: B42627; MUID:92304953; PMID:1610822  
A'Accession: B42627  
A'Molecule type: protein  
A'Results: 1-161 <WAL>  
A'Cross-references: UNIPROT:Q7M3Q2  
C'Comment: This protein becomes a component of the quinine-tanned cement in the tubes by F130.32,53,56,76,89,92,105,108,129,136,149/Modified site: 3',4'-dihydroxyphenylalanine

Query Match  
Best Local Similarity 39.4%; Score 57.5; DB 2; Length 161;  
Matches 13; Conservative 1; Mismatches 10; Indels 19; Gaps 2;

QY 1 WSHPOFEKGGSGG-----GSGW-----SHPOFEK 24

DB 42 WCHPAVHKAAGGCGYGAHPAAGAGGAGGCGGCGYGAHPAVHK 84

RESULT 8  
S35500  
heterogeneous ribonucleic acid particle protein homolog - *Caenorhabditis elegans*  
N'Alternative names: heterogeneous nuclear ribonucleoprotein homolog  
C'Species: *Caenorhabditis elegans*  
C'Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 09-Jul-2004  
C'Accession: S35500; T32620  
R'Iwasaki, M.; Okumura, K.; Kondo, Y.; Tanaka, T.; Igarashi, H.  
Nucleic Acids Res. 20, 4001-4007, 1992  
A>Title: cDNA cloning of a novel heterogeneous nuclear ribonucleoprotein gene homologue  
A'Accession: S35500; MUID:92375684; PMID:1354852  
A'Accession: S35500  
A'Molecule type: mRNA  
A'Results: 1-346 <IWA>  
A'Cross-references: UNIPROT:Q22037; EMBL:S43152  
R'Du, Z.; Scheet, P.; Andrews, S.  
submitted to the EMBL Data Library, December 1997  
A'Description: The sequence of C. elegans cosmid F42A6.  
A'Accession: T32620  
A'Status: preliminary; translated from GB/EMBL/DBJ  
A'Molecule type: DNA  
A'Results: 1-346 <DU>  
A'Cross-references: EMBL:AF038613; PDB:AAK92051.1; GSPDB:GN00022; CESP:F42A6.7  
A'Experimental source: strain Bristol N2; clone F42A6  
C'Genetics:  
A'Gene: CESP:F42A6.7  
A'Map position: 4  
A'Intons: 9/3; 255/1  
C'Superfamily: helix-decussing protein; ribonucleoprotein repeat homology  
F115-90/Domain: ribonucleoprotein repeat homology <RRM1>  
F115-181/Domain: ribonucleoprotein repeat homology <RRM2>

Query Match  
Best Local Similarity 38.7%; Score 56.5; DB 1; Length 346;  
Matches 11; Conservative 2; Mismatches 8; Indels 7; Gaps 1;

QY 1 WSHPOFEK-----GSGGGGWSHPO 21  
DB 271 WCGPQQGGGGGWSHPOGGGGGGGWSHPO 298

RESULT 9  
T20863  
hypothetical protein F13H10.3 - *Caenorhabditis elegans*  
C'Species: *Caenorhabditis elegans*  
C'Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 15-Oct-1999  
C'Accession: T20863  
R'Courtage, A.  
submitted to the EMBL Data Library, January 1996  
A'Accession: T20863  
A'Status: preliminary; translated from GB/EMBL/DBJ  
A'Molecule type: DNA  
A'Results: 1-621 <WIL>  
A'Cross-references: EMBL:Z68748; PDB:CAA92953.1; GSPDB:GN00022; CESP:F13H10.3  
A'Experimental source: clone F13H10  
C'Genetics:  
A'Gene: CESP:F13H10.3  
A'Map position: 4  
A'Intons: 11/3; 45/2; 136/3; 169/2; 222/1; 302/1; 384/1; 417/1; 487/3; 567/2

Query Match  
Best Local Similarity 38.4%; Score 56; DB 2; Length 621;  
Matches 12; Conservative 4; Mismatches 6; Indels 10; Gaps 1;

QY 1 WSHPOFEK-----GSGGGGWSHPO 22  
DB 317 WTTPQWKKQLYDAVSEMEGERTGDDSWSPDK 348

RESULT 10  
 S71954  
 RNA/sedna-binding protein TAF168 - human  
 N/Alternate names: TAF168-binding protein-associated factor TAF168  
 C/Species: Homo sapiens (man)  
 C/Date: 19-Mar-1998 #sequence\_revision 17-Apr-1998 #text\_change 09-Jul-2004  
 C/Accession: S71954  
 R/Bertolotti, A.; Lutz, Y.; Heard, D.J.; Chambon, P.; Tora, L.  
 EMBO J. 15, 5022-5031, 1996  
 A/Title: hTAF168, a novel RNA/sedna-binding protein with homology to the pro-oncoprotein  
 A/Reference number: S71954; MUID:97045110; PMID:8890175  
 A/Accession: S71954  
 A/Molecule type: mRNA  
 A/Residues: 1-589 <BER>  
 A/Cross-references: UNIPROT:Q92804; EMBL:X98893; NID:G1628402; PIDN:CNA67398.1; PID:G162  
 A/Experimental source: HeLa cells  
 C/Genetics:  
 A/Gene: taf68  
 C/Complex: up to 13 TAF168-binding protein-associated factors (TAF168) form together with  
 A/Description: required for activated transcription; binds RNA and single stranded DNA;  
 C/Superfamily: RNA-binding protein, EMS type; ribonucleoprotein repeat homology  
 C/Keywords: RNA binding  
 F/232-307/Domain: ribonucleoprotein repeat homology <RRM>

Query Match 37.0%; Score 54; DB 2; Length 589;  
 Best Local Similarity 75.0%; Pred. No. 26;  
 Matches 9; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 PPFKGGGGGGG 15  
 DB 319 PPFKGGGGGGG 330

RESULT 11  
 E48423  
 homeotic protein engrailed 2 - human  
 N/Alternate names: homeotic protein En-2  
 C/Species: Homo sapiens (man)  
 C/Date: 31-Dec-1993 #sequence\_revision 31-Dec-1993 #text\_change 16-Aug-2004  
 C/Accession: E48423; A30141  
 R/Ilogan, C.; Hanks, M.C.; Noble-Topham, S.; Nallanathan, D.; Provart, N.J.; Joyner, A.L.  
 Dev. Genet. 13, 345-358, 1992  
 A/Title: Cloning and sequence comparison of the mouse, human, and chicken engrailed gene  
 A/Reference number: A48423; MUID:93185339; PMID:1363401  
 A/Accession: E48423  
 A/Status: preliminary; not compared with conceptual translation  
 A/Molecule type: DNA  
 A/Residues: 1-333 <LOG>  
 A/Cross-references: UNIPROT:P19622  
 R/Pool, S.U.; Law, M.L.; Kao, F.T.; Lau, Y.F.  
 Genomics 4, 225-231, 1989  
 A/Title: Isolation and chromosomal localization of the human En-2 gene.  
 A/Reference number: A30141; MUID:89233109; PMID:265873  
 A/Accession: A30141  
 A/Molecule type: DNA  
 A/Residues: 230-333 <POO>  
 C/Genetics:  
 A/Gene: En-2  
 C/Superfamily: homeobox homology  
 C/Keywords: DNA binding; homeobox; nucleus; transcription regulation  
 F/245-301/Domain: homeobox homology <HOX>

Query Match 37.0%; Score 54; DB 2; Length 333;  
 Best Local Similarity 58.8%; Pred. No. 19;  
 Matches 10; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

QY 4 PPFKGGGGGGGWSHP 20  
 DB 21 PPFKGGGGGGGSSSP 37

RESULT 12  
 F81678  
 GTP-binding protein HflX TC0658 [imported] - Chlamydia muridarum (strain N199)  
 C/Species: Chlamydia muridarum, Chlamydia trachomatis MOPn  
 C/Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004  
 C/Accession: F81678  
 R/Read, T.D.; Brunham, R.C.; Shen, C.; Gill, S.R.; Heidelberg, J.F.; White, O.; Hickey,  
 Nucleic Acids Res. 28, 1397-1406, 2000  
 A/Title: Genome sequences of Chlamydia trachomatis MOPn and Chlamydia pneumoniae AR39.  
 A/Reference number: A81500; MUID:20150255; PMID:10684935  
 A/Accession: F81678  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-447 <TET>  
 A/Cross-references: UNIPROT:Q9PK15; GB:AE002334; GB:AE002160; NID:G7190690; PIDN:AAF3948  
 A/Experimental source: strain N199 (MOPn)  
 C/Genetics:  
 A/Gene: TC0658  
 C/Superfamily: GTP-binding protein hflX; translation elongation factor Tu homology

Query Match 37.0%; Score 54; DB 2; Length 447;  
 Best Local Similarity 60.0%; Pred. No. 26;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 WSHPOFKGGGGGGG 15  
 DB 165 WGHLSRQKSGGGGG 179

RESULT 13  
 G71521  
 probable GTP binding protein - Chlamydia trachomatis (serotype D, strain UW3/Cx)  
 C/Species: Chlamydia trachomatis  
 C/Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 09-Jul-2004  
 C/Accession: G71521  
 R/Stephens, R.S.; Kaiman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell,  
 Science 282, 754-759, 1998  
 A/Title: Genome sequence of an obligate intracellular pathogen of humans: Chlamydia trachomatis  
 A/Reference number: A71570; MUID:9900809; PMID:9784136  
 A/Accession: G71521  
 A/Status: preliminary  
 A/Molecule type: DNA  
 A/Residues: 1-447 <ARN>  
 A/Cross-references: UNIPROT:O84183; GB:AE001311; GB:AE001273; NID:G3328799; PIDN:AA6797  
 A/Experimental source: serotype D, strain UW-3/Cx  
 C/Genetics:  
 A/Gene: hflX  
 C/Superfamily: GTP-binding protein hflX; translation elongation factor Tu homology  
 C/Keywords: nucleotide binding; P-loop  
 F/233-240/Region: nucleotide-binding motif A (P-loop)  
 F/346-349/Region: GTP-binding NKXD motif  
 F/371-373/Region: GTP-binding SAK/L motif

Query Match 37.0%; Score 54; DB 2; Length 447;  
 Best Local Similarity 60.0%; Pred. No. 26;  
 Matches 9; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 WSHPOFKGGGGGGG 15  
 DB 165 WGHLSRQKSGGGGG 179

RESULT 14  
 T08420  
 1-phosphatidylinositol 3-kinase (EC 2.7.1.137) - Chlamydomonas reinhardtii (fragment)  
 N/Alternate names: phosphatidylinositol 3-kinase  
 C/Species: Chlamydomonas reinhardtii  
 C/Date: 11-Jun-1999 #sequence\_revision 11-Jun-1999 #text\_change 09-Jul-2004  
 C/Accession: T08420  
 R/Molendijk, A.J.; Irvine, R.F.  
 Plant Mol. Biol. 37, 53-66, 1998



A>Title: Inositolide signalling in Chlamydomonas: Characterization of a phosphatidylinositol  
 A:Reference number: Z16411; PMID:98281574; PMID:9620264  
 A:Accession: T08420  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-732 <MOL>  
 A:Cross-references: UNIPROT:O04269; EMBL:U97662; NID:G2109288; PIDN:AAC50017.1; PID:G210  
 A:Experimental source: strain cw-15  
 C:Function:  
 A:Description: catalyzes the phosphorylation of phosphoinositides on the 3-hydroxyl group  
 C:Superfamily: phosphatidylinositol 3-kinase, Vps34 type  
 C:Keywords: phosphotransferase; signal transduction

Query Match 37.0%; Score 54; DB 2; Length 732;  
 Best Local Similarity 50.0%; Pred. No. 42;  
 Matches 9; Conservative 0; Mismatches 9; Indels 0; Gaps 0;

QY 3 HPOFEKGGGGGSGWSHP 20  
 DB 379 HSPFGGGGGGGGWPP 396

## RESULT 15

E96603  
 unknown protein F14G9.26 [imported] - Arabidopsis thaliana  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004  
 C:Accession: E96603  
 R:Theologis, A.; Eckert, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso,  
 Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;  
 ansen, N.F.; Hughes, B.; Hultzer, L.  
 Nature 408, 816-820, 2000  
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.  
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani,  
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.  
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shim, P.; Southwick, A.M.; Sun, H.; Tallon,  
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.  
 A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.  
 A:Reference number: A86141; MIMD:21016719; PMID:11130712  
 A:Accession: E96603  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-154 <STO>  
 A:Cross-references: UNIPROT:Q9C7J5; GB:AE005173; NID:G11094725; PIDN:AAG29660.1; GSPDB:G  
 C:Genetics:  
 A:Gene: F14G9.26  
 A:Map position: 1

Query Match 35.6%; Score 52; DB 2; Length 154;  
 Best Local Similarity 80.0%; Pred. No. 16;  
 Matches 8; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 9 GGSGGGGWS 18  
 DB 115 GGSGGGGWS 124

Search completed: March 2, 2005, 12:28:57  
 Job time : 19.3415 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 69.1707 Seconds  
(without alignments)  
137.824 Million cell updates/sec

Title: SEQ11  
Perfect score: 146  
Sequence: 1 wshpofekyggsgggswshpofek 24

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues  
Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt 03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	64	43.8	163	061UR6	061UR6 drosophila
2	62	42.5	429	062N21	062N21 burkholderi
3	62	42.5	429	063XB7	063XB7 burkholderi
4	62	42.5	704	08CUD2	08CUD2 rattus norv
5	61	41.8	99	07DL91	07DL91 anabaena va
6	61	41.8	99	09WX37	09WX37 anabaena ap
7	61	41.8	743	08CH90	08CH90 rattus norv
8	61	41.8	929	09LGE7	09LGE7 oryza sativ
9	60	41.1	448	1 P031_HUMAN	003052 homo sapien
10	60	41.1	454	09VHK9	09VHK9 drosophila
11	60	41.1	455	0960C1	0960C1 drosophila
12	60	41.1	809	07OG12	07OG12 anopheles g
13	58.5	40.1	82	07SKA2	07SKA2 oryza sativ
14	58	39.7	150	07X243	07X243 citrobacter
15	58	39.7	449	1 P031_MOUSE	F21952 mus musculu
16	58	39.7	451	1 P031_RAT	P20367 rattus norv
17	58	39.7	690	07CS58	07CS58 agrobacteri
18	58	39.7	1078	1 CARB_BUCAI	P57344 buchnera ap
19	57.5	39.4	161	2 07M3Q2	07M3Q2 phytomyces
20	57	39.0	1198	2 07S153	07S153 neurospora
21	57	39.0	1515	2 06AV56	06AV56 oryza sativ
22	56.5	38.7	309	2 09S569	09S569 caenorhabd
23	56.5	38.7	346	2 022037	022037 caenorhabd
24	56	38.4	259	1 PRIO_TRIVU	P51780 trichosurus
25	56	38.4	434	2 08L4Y9	08L4Y9 oryza sativ
26	56	38.4	517	2 06ADK6	06ADK6 leishmania x
27	56	38.4	615	1 YSRK_CAEEL	Q19425 caenorhabd
28	56	38.4	617	2 08T3E2	08T3E2 caenorhabd
29	56	38.4	640	2 08A120	08A120 bacteroides
30	56	38.4	718	2 07XAP4	07XAP4 oryza sativ
31	56	38.4	808	2 06BYP9	06byp9 debaryomyce

32	55.5	38.0	285	2	06H7B8	06H7B8 oryza sativ
33	55.5	38.0	420	2	06RR45	06RR45 arabidopsis
34	55	37.7	204	2	09ACP9	09ACP9 oryza sativ
35	55	37.7	232	2	07PSQ8	07PSQ8 anopheles g
36	55	37.7	397	2	095XG8	095XG8 macaca fasc
37	55	37.7	501	2	086X94	086X94 homo sapien
38	55	37.7	557	2	08B046	08B046 mus musculu
39	55	37.7	592	1	R856_HUMAN	092804 homo sapien
40	55	37.7	959	2	06ZDR8	06ZDR8 oryza sativ
41	54.5	37.3	415	2	07PST4	07PST4 anopheles g
42	54	37.0	106	2	08LTI1	08LTI1 lycopersico
43	54	37.0	265	2	06H6P3	06H6P3 oryza sativ
44	54	37.0	317	2	094LW4	094LW4 oryza sativ
45	54	37.0	333	1	HME2_HUMAN	P19622 homo sapien

## ALIGNMENTS

RESULT 1	061UR6	PRELIMINARY;	PRT;	163 AA.
ID	061UR6			
AC	061UR6			
DT	05-JUL-2004 (TrEMBLrel. 27, Created)			
DT	05-JUL-2004 (TrEMBLrel. 27, Last sequence update)			
DT	05-JUL-2004 (TrEMBLrel. 27, Last annotation update)			
DE	HDCL4384.			
GN	ORFNames=HDCL4384;			
OS	Drosophila melanogaster (Fruit fly).			
OC	Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;			
OC	Neoptera; Endopterygota; Diptera; Brachyera; Muscomorpha;			
OC	Ephydroidea; Drosophilidae; Drosophila.			
OX	NCBI_TaxID=7227;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RX	PubMed=14709175; DOI=10.1186/gb-2003-5-1-r3;			
RA	Hild M., Beckmann B., Haas S., Koch B., Solovjev V., Busold C.,			
RA	Fellenberg K., Boutros M., Vingron M., Sauer F., Hobeisel J., Pao R.,			
RT	"An integrated gene annotation and transcriptional profiling approach			
RT	towards the full gene content of the Drosophila genome.";			
RL	Genome Biol. 5:R3-R3(2003).			
CC	-1- MISCELLANEOUS: The sequence shown here is derived from an			
CC	EMBL/GenBank/DBJ third party annotation (TPA) entry.			
DR	EMBL, BK002650; DAA04156.1; -- 002B716DAAABBD39 CRC64;			
SQ	SEQUENCE 163 AA; 18316 MW; 002B716DAAABBD39 CRC64;			
Query Match	43.8%; Score 64; DB 2; Length 163;			
Best Local Similarity	66.7%; Pred. No. 3.8;			
Matches	10; Conservative 1; Mismatches 4; Indels 0; Gaps 0;			
OY	1 WSHPOFEKGGSGGG 15			
Db	77 WPHPOQPRGGGGGGG 91			
RESULT 2	062N21	PRELIMINARY;	PRT;	429 AA.
ID	062N21			
AC	062N21			
DT	25-OCT-2004 (TrEMBLrel. 28, Created)			
DT	25-OCT-2004 (TrEMBLrel. 28, Last sequence update)			
DT	25-OCT-2004 (TrEMBLrel. 28, Last annotation update)			
DE	Capuslar polysaccharide biosynthesis/export periplasmic protein.			
GN	ORFNames=BXA0047;			
OS	Burkholderia mallei ATCC 23344.			
OC	Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;			
OC	Burkholderiaceae; Burkholderia.			
OX	NCBI_TaxID=243160;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=ATCC 23344;			
RA	Nierman W.C., Deshaizer D., Kim H.S., Tettelin H., Nelson K.E.,			
RA	Feldblyum T., Ulrich R.L., Ronning C.M., Brinkac L.M., Daugherty S.C.,			

RA Davidson T.D., Deboy R.T., Dimitrov G., Dodson R.J., Durkin A.S.,  
 RA Gwin M.L., Haft D.H., Khouri H., Kolonay J.P., Madupu R.,  
 RA Mohammad Y., Nelson W.C., Radune D., Romero C.M., Saria S.,  
 RA Selengut J., Shambin C., Sullivan S.A., White O., Yu Y., Zafar N.,  
 RA Zhou L., Fraser C.M.,  
 RT "Structural flexibility in the Burkholderia mallei genome."  
 RL Proc. Natl. Acad. Sci. U.S.A. 101:14247-14251(2004).  
 DR EMBL; CP000010; AAU48949.1; -.  
 SO SEQUENCE 429 AA; 45905 MW; 614129D67FBED97 CRC64;

Query Match 42.5%; Score 62; DB 2; Length 429;  
 Best Local Similarity 45.0%; Pred. No. 19;  
 Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 WSHPOFKGGSGGSGWSMP 20  
 DB 96 WDHPELTRGGNGTGADTSP 115

## RESULT 3

AC 063XB7 PRELIMINARY; PRT; 429 AA.  
 DT 25-OCT-2004 (TRENBLrel. 28, Created)  
 DT 25-OCT-2004 (TRENBLrel. 28, Last sequence update)  
 DE Putative polyaccharide biosynthesis/export protein.  
 GN ORFNames=BPSL0619.  
 OS Burkholderia pseudomallei K96243.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Burkholderia.  
 OX NCBI\_TaxID=272560;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=K96243;  
 RX PubMed=15377794;  
 RA Holden M.T.G., Tibball R.W., Peacock S.J., Cerdano-Tarraga A.M.,  
 RA Ackins T., Crossman L.C., Platt T., Churcher C., Mungall K.,  
 RA Bentley S.D., Sebahia M., Thomson N.R., Baason N., Beacham I.R.,  
 RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,  
 RA Chillingworth T., Cronin A., Crosser B., Davis P., Desnazer D.,  
 RA Fellwell T., Fraser A., Hance Z., Hauser H., Holtroyd S., Jagels K.,  
 RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,  
 RA Rabinowitch E., Rutherford K., Sanders M., Simmonds M.,  
 RA Songvilay S., Stevens K., Tunapa S., Vesaratchavest M.,  
 RA Whitehead S., Yeates C., Barrett B.G., Oyston P.C.F., Parkhill J.,  
 RT "Genomic plasticity of the causative agent of melioidosis."  
 RT Burkholderia pseudomallei.  
 RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245(2004).  
 DR EMBL; BX571965; CAH34612.1; -.  
 SO SEQUENCE 429 AA; 45877 MW; 7A4129C970AECC98 CRC64;

Query Match 42.5%; Score 62; DB 2; Length 429;  
 Best Local Similarity 45.0%; Pred. No. 19;  
 Matches 9; Conservative 5; Mismatches 6; Indels 0; Gaps 0;

QY 1 WSHPOFKGGSGGSGWSMP 20  
 DB 96 WDHPELTRGGNGTGADTSP 115

## RESULT 4

AC 08CUD2 PRELIMINARY; PRT; 704 AA.  
 DT 01-MAR-2003 (TRENBLrel. 23, Created)  
 DT 01-MAR-2003 (TRENBLrel. 23, Last sequence update)  
 DE Guanylyl cyclase alpha 1 subunit.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nakamura I., Suzuki N.,  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB096020; BAC24016.1; -.  
 DR HSSP; P30803; IAZS.  
 DR GO; GO:0004383; F:guanylate cyclase activity; IEA.  
 DR GO; GO:0016829; F:intracellular signaling cascade; IEA.  
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.  
 DR InterPro; IPR001054; G\_cyclase.  
 DR Pfam; PF00211; Guanylate\_cyc; 1.  
 DR SMART; SM00044; CYC; 1.  
 DR PROSITE; PS00452; GUANYLATE\_CYCLASES\_1; 1.  
 DR PROSITE; PS50125; GUANYLATE\_CYCLASES\_2; 1.  
 KW lyase.  
 SO SEQUENCE 704 AA; 79188 MW; F73FC97B68544E6A CRC64;

Query Match 42.5%; Score 62; DB 2; Length 704;  
 Best Local Similarity 41.0%; Pred. No. 32;  
 Matches 16; Conservative 2; Mismatches 5; Indels 16; Gaps 2;

QY 2 SHPOFK-----GGSG-----GSGWSHPQFK 24  
 DB 666 SKFWFOQDAEDGNANFLGKASGVDLVRCGWSHPOFK 704

## RESULT 5

AC Q7DL91 PRELIMINARY; PRT; 99 AA.  
 ID Q7DL91  
 DT 05-JUL-2004 (TRENBLrel. 27, Created)  
 DT 05-JUL-2004 (TRENBLrel. 27, Last sequence update)  
 DE RNA binding protein.  
 GN Name=rbpE;  
 OS Anabaena variabilis.  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Anabaena.  
 OX NCBI\_TaxID=1172;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=WM3;  
 RX MEDLINE=99216503; PubMed=10198437; DOI=10.1093/nar/27.9.2029;  
 RA Maruyama K., Sato N., Ohta N.,  
 RT "Conservation of structure and cold-regulation of RNA-binding proteins  
 in cyanobacteria; probable convergent evolution with eukaryotic  
 glycine-rich RNA-binding proteins."  
 RT Nucleic Acids Res. 27:2029-2036(1999).  
 RL EMBL; AB003331; BAA77711.1; -.  
 DR InterPro; IPR000504; RNA\_rec\_mot.  
 DR Pfam; PF00076; RRM\_1; 1.  
 DR SMART; SM00360; RRM; 1.  
 DR PROSITE; PS50102; RRM; 1.  
 DR PROSITE; PS00030; RNP\_RNP\_1; UNKNOWN 1.  
 SO SEQUENCE 99 AA; 10811 MW; 893B196FC8187798 CRC64;

Query Match 41.8%; Score 61; DB 2; Length 99;  
 Best Local Similarity 64.7%; Pred. No. 5.1;  
 Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 2 SHPOFKGGSGGSGWS 18  
 DB 77 ARPREFKARGSGGWS 93

## RESULT 6

AC 09WX37 PRELIMINARY; PRT; 99 AA.  
 ID 09WX37  
 DT 01-NOV-1999 (TRENBLrel. 12, Created)  
 DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)  
 DE RNA binding protein.  
 GN Name=rbpE; OrderedLocuNames=a112777;

OS Anabaena sp. (strain PCC 7120).  
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.  
 OX NCBI\_TaxId=103690;  
 RN  
 ID  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=21595285; PubMed=11759840;  
 RA Kaneko T., Nakamura Y., Molk C.P., Kuritz T., Sasamoto S.,  
 RA Watanabe A., Iriuguchi M., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,  
 RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,  
 RA Yasuda M., Tabata S.,  
 RT "Complete genomic sequence of the filamentous nitrogen-fixing  
 cyanobacterium Anabaena sp. strain PCC 7120."  
 RL DNA Ref. 8:205-213(2001).  
 DR EMBL, AP003590; BAB74476.1; -.  
 DR PIR, AB2153; AB2153.  
 DR HSSP, P33240; 1P1T.  
 DR InterPro, IPR000504; RNA\_rec\_mot.  
 DR Pfam, PF00076; RRM\_1; 1.  
 DR SMART, SM00360; RRM; 1.  
 DR PROSITE, PS50102; RRM; 1.  
 DR PROSITE, PS00030; RRM\_RNP\_1; 1.  
 KM Complete proteome.  
 SQ SEQUENCE 99 AA; 10811 MW; 893B196FC8187798 CRC64;

Query Match 41.8%; Score 61; DB 2; Length 99;  
 Best Local Similarity 64.7%; Pred. No. 5.1;  
 Matches 11; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Oy 2 SHPOFEKGGSGGSGWS 18  
 Db 77 ARPREKGRSGGSGWS 93

RESULT 7  
 OQ8CH90 PRELIMINARY; PRT; 743 AA.

AC OQ8CH90;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DE 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 OS Soluble guanylyl cyclase alpha 2 subunit E219G mutant.  
 OC Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN  
 ID  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RA Nakamura I., Yao Y., Suzuki N.;  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AB097860; BAC44887.1; -.  
 DR HSSP, P30803; 1A2S.  
 DR GO, GO:0004383; F:guanylate cyclase activity; IEA.  
 DR GO, GO:0016829; F:lipase activity; IEA.  
 DR GO, GO:0007242; P:intracellular signaling cascade; IEA.  
 DR InterPro, IPR001054; G\_cyclase.  
 DR InterPro, IPR009080; tRNA\_Nasyl\_a\_bind.  
 DR Pfam, PF00211; Guanylate\_cyc; 1.  
 DR SMART, SM00442; CYCC; 1.  
 DR PROSITE, PS00454; GUANYLATE\_CYCLASPS\_1; 1.  
 DR PROSITE, PS50125; GUANYLATE\_CYCLASPS\_2; 1.  
 KM Lyase.  
 SQ SEQUENCE 743 AA; 83251 MW; ACF5C53E0982813A CRC64;

Query Match 41.8%; Score 61; DB 2; Length 743;  
 Best Local Similarity 100.0%; Pred. No. 44;  
 Matches 10; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Oy 15 GSMWHPQFEK 24  
 Db 734 GSMWHPQFEK 743

RESULT 8  
 OQ9LGE7 PRELIMINARY; PRT; 929 AA.  
 AC OQ9LGE7;  
 DT 01-OCT-2000 (TREMBLrel. 15, Created)  
 DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)  
 DE 25-OCT-2004 (TREMBLrel. 28, Last annotation update)  
 DE P0406H10.1 protein (P0509B06.3 protein).  
 GN Name=P0406H10.1; Synonym=P0509B06.3;  
 OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliopsida; Liliopsida; Poales; Poaceae;  
 OC Eriarchoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxId=39947;  
 RN  
 ID  
 RP SEQUENCE FROM N.A.  
 RX PubMed=12447438; DOI=10.1038/nature01184;  
 RA Sasaki T., Matsumoto T., Yamamoto K., Sakata K., Baba T., Karayose Y.,  
 RA Wu J., Nimura Y., Cheng Z., Nagamura Y., Antonio B.A., Kanamori H.,  
 RA Hosokawa S., Masukawa M., Arikawa K., Chiden Y., Hayashi M.,  
 RA Okamoto M., Ando T., Aoki H., Arita K., Hamada M., Harada C.,  
 RA Hijiishi S., Honda M., Ichikawa Y., Idonuma A., Iijima M., Ikeda M.,  
 RA Ikono M., Itoh S., Itoh T., Itoh Y., Iwabuchi A., Kamiya K.,  
 RA Karasawa W., Katagiri S., Kikuta A., Kobayashi N., Kono I.,  
 RA Machita K., Maehara T., Mizuno H., Mizubayashi T., Mukai Y.,  
 RA Nagasaki H., Nakashima M., Nakama Y., Nakamichi Y., Nakamura M.,  
 RA Namiki N., Negishi M., Ohta I., Ono N., Saji S., Sakai K., Shibata M.,  
 RA Shimokawa T., Shomura A., Song J., Takazaki Y., Terasawa K., Tsuji K.,  
 RA Waki K., Yamagata H., Yamane H., Yoshiki S., Yoshihara R., Yokawa K.,  
 RA Zhong H., Iwama H., Endo T., Ito H., Hahn J.H., Kim H.I., Eun M.Y.,  
 RA Yano M., Jiang J., Gojobori T.,  
 RT "The genome sequence and structure of rice chromosome 1."  
 RL Nature 420:312-316(2002).  
 DR EMBL, AP002524; BAB07977.1; -.  
 DR EMBL, AP002903; BAB63468.1; -.  
 DR Gramene; O9LGE7; -.  
 DR InterPro, IPR002794; DUF92.  
 DR Pfam, PF01940; DUF92; 1.  
 DR Prodom; PD014594; DUF92; 1.  
 SQ SEQUENCE 929 AA; 100025 MW; F5869AD16F3F6A78 CRC64;

Query Match 41.8%; Score 61; DB 2; Length 929;  
 Best Local Similarity 52.9%; Pred. No. 56;  
 Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Oy 1 WSHPOFEKGGSGGWS 17  
 Db 537 WSPAMDGGGGGGVW 553

RESULT 9  
 ID P031 HUMAN STANDARD; PRT; 448 AA.

AC P031 HUMAN;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE POU domain, class 3, transcription factor 1 (Octamer-binding  
 DE transcription factor 6) (Oct-6) (POU-domain transcription factor  
 DE SCIP).  
 GN Name=POU3F1; Synonym=OCT6, OTF6;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxId=9606;  
 RN  
 ID  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=94217723; PubMed=7909356;  
 RA Faus I., Heu H.J., Fuchs E.,  
 RT "Oct-6: a regulator of keratinocyte gene expression in stratified  
 skin squamous epithelia."  
 RL Mol. Cell. Biol. 14:3263-3275(1994).

```

RN [2]
RP SEQUENCE OF 51-448 FROM N.A.
RX MEDLINE=93197134; PubMed=8451175;
RA Tobler A., Schreiber E., Fontana A.;
RT "The human Oct-6 POU transcription factor lacks the first 50 amino
RL Nucleic Acids Res. 21:1043-1043(1993).
CC -1- FUNCTION: Transcription factor that binds to the octamer motif
CC (5'-ATTTGAT-3'). Thought to be involved in early embryogenesis
CC and neurogenesis.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: Expressed in embryonal stem cells and in the
CC developing brain.
CC -1- SIMILARITY: Belongs to the POU transcription factor family. Class-
CC 3 subfamily.
CC -1- SIMILARITY: Contains 1 homeobox domain.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; L26494; AA59965.1; -
CC EMBL; Z18284; CAA79158.1; -
CC PIR; A56018; A56018.
CC HSSP; P14859; 1CQT.
CC DR TRANSFAC; T00655; -
CC DR Genew; HGNC:9214; POU3FL.
CC DR MIM; 602479; -
CC DR GO; GO:0003700; F:transcription factor activity; TMS.
CC DR GO; GO:0008366; P:nervous ensheathment; TMS.
CC DR InterPro; IPR001356; Homeobox.
CC DR InterPro; IPR009057; Homeodomain like.
CC DR InterPro; IPR010982; Lambda_like_DNA.
CC DR InterPro; IPR000387; POU.
CC DR InterPro; IPR007103; POU_homeo.
CC DR Pfam; PF00046; Homeobox; 1.
CC DR Pfam; PF00157; POU; 1.
CC DR PRINTS; PR00024; HOMEBOX.
CC DR PRINTS; PR00028; POUHOMEIN.
CC DR PRODOM; PD000583; POU; 1.
CC DR PROSITE; PS00027; HOMEBOX_1; 1.
CC DR PROSITE; PS00071; HOMEBOX_2; 1.
CC DR PROSITE; PS00035; POU_1; 1.
CC DR PROSITE; PS00465; POU_2; 1.
CC DR DNA-binding; Homeobox; Nuclear protein; Transcription regulation.
CC FT DOMAIN 248 318 POU.
CC FT DNA BIND 336 395 Homeobox.
CC SQ SEQUENCE 448 AA; 45270 MW; 3AA45E029A2C8421 CRC64;
QY
QY Query Match 41.1%; Score 60; DB 1; Length 448;
QY Best Local Similarity 45.5%; Pred. No. 34;
QY Matches 10; Conservative 3; Mismatches 9; Indels 0; Gaps 0;
QY 2 SHPOFKGGGGGGGWSHPQFE 23
QY :|||:|||||:
QY AC O9VBK9; PRELIMINARY; PRT; 454 AA.
QY AC O9VBK9;
QY DT 01-MAY-2000 (TrEMBLrel. 13, Created)
QY DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
QY DT 01-OCT-2002 (TrEMBLrel. 22, Last annotation update)
QY DE CG5913-PA.
QY GN ORFNames=CG5913;
QY OS Drosophila melanogaster (Fruit fly).

```

```

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20196006; PubMed=10731132; DOI=10.1126/science.287.5461.2185;
RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA Amanatides P.G., Scherier S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA Brandon R.C., Rogers Y.H., Blazer J.G., Champe M., Pfeiffer B.D.,
RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Gabor G.L.,
RA Abail J.F., Agbayani A., An H.J., Andrews-Pfannkoch C., Baldwin D.,
RA Ballew R.M., Baau A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA Beeson K.Y., Benos P.V., Bereman B.P., Bhandari D., Bolshakov S.,
RA Borokova D., Botchan M.R., Bouck J., Brockstein P., Brotler P.,
RA Buttis K.C., Buesam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P.,
RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA Duthin K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
RA Foster C., Gong F., Gorrell J.H., Gu Z., Guan P., Harris K.,
RA Harris N.L., Harvey D., Heitman T.J., Hernandez J.R., Houck J.,
RA Hostin D., Houston K.A., Howland T.J., Wei M.H., Ibegwan C.,
RA Jaitai M., Kalush F., Karen G.H., Ke Z., Kemison J.A., Ketchum K.A.,
RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA Laeko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA Nelson D.R., Nelson K.A., Nixon K., Pollard J., Puri V., Reese M.G.,
RA Palazzolo M., Peltman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA Reinert K., Remington K., Saunders R.D., Scheeler F., Shen H.,
RA Shie B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA Spler E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA Svitek R., Tector C., Turner C., Turner R., Venter E., Wang A.H., Wang X.,
RA Wang Z.Y., Wasarman D.A., Weinstein G.M., Weissbach J.,
RA Williams S.M., Woodger, Morley K.C., Wu D., Yang S., Yao Q.A., Ye J.,
RA Yeh R.F., Zaveri J.S., Zhao M., Zhang G., Zhu Q., Zheng L.,
RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu X., Zhu X., Smith H.O.,
RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT "The genome sequence of Drosophila melanogaster."
RT Science 287:2185-2195(2000).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426065; PubMed=12537568;
RA Celniker S.E., Wheeler D.A., Krommiller B., Carlson J.W., Halpern A.,
RA Patel S., Adams M., Champe M., Dugan S.P., Frise E., Hodgson A.,
RA George R.A., Hoskins R.A., Laverly T., Muzny D.M., Nelson C.R.,
RA Paclele J.M., Park S., Pfeiffer B.D., Richards S., Sodergren E.J.,
RA Svitek R., Taber F.E., Wan K., Stapleton M., Sutton G.G., Venter C.,
RA Weinstein G., Scherer S.E., Myers E.W., Gibbs R.A., Rubin G.M.;
RT "Finishing a whole-genome shotgun: Release 3 of the Drosophila
RT melanogaster euchromatic genome sequence."
RT Genome Biol. 3:RESEARCH0079-RESEARCH0079(2002).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426070; PubMed=12537573;
RA Kaminker J.S., Bergman C.M., Krommiller B., Carlson J., Svitek R.,
RA Patel S., Frise E., Wheeler D.A., Lewis S.E., Rubin G.M.,
RA Ashburner M., Celniker S.E.;
RT "The transposable elements of the Drosophila melanogaster euchromatin:
RT a genomic perspective."
RT Genome Biol. 3:RESEARCH0084-RESEARCH0084(2002).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=22426069; PubMed=12537572;
RA Miera S., Crosby M.A., Mungall C.J., Matthews B.B., Campbell K.S.,
RA Hradecky P., Huang Y., Kaminker J.S., Milburn G.H., Prochuk S.E.,
RA Smith C.D., Tupy J.L., Whitfield E.J., Bayraktaroglu L., Berman B.P.,
RA Betencourt B.R., Celniker S.E., de Grey A.D., Drysdale R.A.

```

RA Harris N.L., Richter J., Russo S., Schroeder A.J., Shu S.O.,  
 RA Starlecon M., Yamada C., Ashburner M., Gelbart W.M., Rubin G.M.,  
 RA Lewis S.E.;  
 "Annotation of the *Drosophila melanogaster* euchromatic genome: a  
 RT systematic review."  
 RL Genome Biol. 3:RESEARCH0083-RESEARCH0083(2002).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RG FlyBase)  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
 RP SEQUENCE FROM N.A.  
 RG FlyBase)  
 RL Submitted (MAR-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AE00754; AAF56525.1; --  
 DR Intact; Q9YBK9; --  
 DR FlyBase; FBgn0039385; CG5913.  
 SQ SEQUENCE 454 AA; 49967 MW; 893BA66A5E46CD17 CRC64;

Query Match 41.1%; Score 60; DB 2; Length 454;  
 Best Local Similarity 69.2%; Pred. No. 35;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 9 GGGGGGGGWSHPQ 21  
 DB 364 GGGGGGGGWSHPQ 376

RESULT 11  
 Q960C1 PRELIMINARY; PRT; 455 AA.

DT 01-DEC-2001 (TREMBLrel. 19, Created)  
 DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)  
 DT 01-OCT-2002 (TREMBLrel. 22, Last annotation update)  
 DE SD08037P.  
 GN ORFName=CG5913;  
 OS *Drosophila melanogaster* (Fruit fly).  
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;  
 OC Ephydroidea; Drosophilidae; *Drosophila*.  
 OX NCBI\_TaxID=7227;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Starlecon M., Brokstein P., Hong L., Agbayani A., Carlson J.,  
 RA Champagne M., Chavez C., Dorsett V., Farfan D., Friese B., George R.,  
 RA Gonzalez M., Guatin H., Li P., Liao G., Miranda A., Mungall C.J.,  
 RA Nunco J., Paclieb J., Paragae V., Park S., Phoumenavong S., Wan K.,  
 RA Yu C., Lewis S.E., Rubin G.M., Celisner S.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL, AY052130; AK93554.1; --  
 DR FlyBase; FBgn0039385; CG5913.  
 SQ SEQUENCE 455 AA; 50095 MW; ED6A9A81669D444D CRC64;

Query Match 41.1%; Score 60; DB 2; Length 455;  
 Best Local Similarity 69.2%; Pred. No. 35;  
 Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 9 GGGGGGGGWSHPQ 21  
 DB 364 GGGGGGGGWSHPQ 376

RESULT 12

Q70G12 PRELIMINARY; PRT; 809 AA.  
 AC Q70G12;  
 DT 01-MAR-2004 (TREMBLrel. 26, Created)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE AGCP13861 (Fragment).  
 GN Name=agc51770; ORFName=ENSGANG0000012651;  
 OS *Anopheles gambiae* str. PEST.

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;  
 OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidae; Anopheles.  
 OX NCBI\_TaxID=180454;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RG STRAIN=PEST;  
 RA Anopheles Genome Sequencing Consortium;  
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.  
 CC -1- CAUTION: The sequence shown here is derived from an  
 EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 preliminary data.  
 CC EMBL; AAB01008844; EAA05947.1; --  
 DR InterPro; IPR009818; Ataxin-2\_C.  
 DR InterPro; IPR009604; Ataxin-2\_N.  
 DR InterPro; IPR010920; Sm\_like\_fiboprot.  
 DR Pfam; PF06741; Ataxin-2\_N; 1.  
 DR Pfam; PF07145; PAM2; 1.  
 FT NON TER 1  
 SQ SEQUENCE 809 AA; 86737 MW; 636706D304EFDDBD CRC64;

Query Match 41.1%; Score 60; DB 2; Length 809;  
 Best Local Similarity 56.2%; Pred. No. 64;  
 Matches 9; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 7 EKGGGGGGGWSHPQ 22  
 DB 243 QHGGAGGGGGGHPQY 258

RESULT 13  
 Q75KA2 PRELIMINARY; PRT; 82 AA.

DT 05-JUL-2004 (TREMBLrel. 27, Created)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
 DE Hypothetical protein OSJNB0106M04.14.  
 GN Name=OSJNB0106M04.14;  
 OS *Oryza sativa* (Japanese cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzoideae; *Oryza*.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Buehl C.R., Yuan Q., Ouyang S., Liu J., Gansberger K., Jones K.M.,  
 RA Overton II L.L., Tsaltrin T., Kim M.M., Bera J.U., Jin S.S.,  
 RA Padrosh D.W., Tallon L.J., Koo H., Zismann V., Hsiao J., Blunt S.,  
 RA Vanden S.S., Riedmiller S.B., Uteback T.T., Feldlyum T.V.,  
 RA Yang Q.Q., Haas B.J., Suh B.B., Peterson J.O., Quackenbush J.,  
 RA White O., Salzberg S.L., Fraser C.M.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Buehl R.;  
 RL Submitted (JAN-2004) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AC107207; AAR87333.1; --  
 KW Hypothetical protein.  
 SQ SEQUENCE 82 AA; 9050 MW; 52B311BC304987C9 CRC64;

Query Match 40.1%; Score 58.5; DB 2; Length 82;  
 Best Local Similarity 40.7%; Pred. No. 8.4;  
 Matches 11; Conservative 2; Mismatches 5; Indels 9; Gaps 1;

QY 4 POFKGGGGGWSHPQ 21  
 DB 14 PELSGGGGGGGRGRRRRRMSR 40

RESULT 14  
 Q7X243 PRELIMINARY; PRT; 150 AA.  
 ID Q7X243;  
 AC Q7X243;





GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 108.098 Seconds  
(without alignments)  
85.869 Million cell updates/sec

Title: SEQ11  
Perfect score: 146  
Sequence: 1 wshpqfkygsgsggawshpqfky 24

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	130	89.0	36	ABP60370	Abp60370 Streptavi
2	84	57.5	396	AAW93966	Aaw93966 Plasmid p
3	84	57.5	396	AAAB46420	Aab46420 Bilin bin
4	84	57.5	396	AAAB46423	Aab46423 Bilin bin
5	84	57.5	400	ADA27291	Ada27291 Plasmid p
6	84	57.5	400	ADA27292	Ada27292 Plasmid p
7	84	57.5	400	ADA00700	Ada00700 Modified
8	84	57.5	400	ADA00707	Ada00707 Modified
9	84	57.5	400	ADA00701	Ada00701 Modified
10	75	51.4	659	AAAB46427	Aab46427 Bilin bin
11	71.5	49.0	485	ADS20249	Ads20249 Furin-pro
12	71.5	49.0	697	ADS20227	Ads20227 Human agg
13	69.5	47.6	845	ADS20243	Ads20243 Human agg
14	69	47.3	24	ABP60362	Abp60362 Streptavi
15	66	45.2	42	AAOI1986	Aaoi1986 Protein o
16	63.5	43.5	35	ABP60369	Abp60369 Streptavi
17	63	43.2	42	AAOI1988	Aaoi1988 Protein o
18	63	43.2	183	ADA27294	Ada27294 Plasmid p
19	63	43.2	183	ADA00703	Ada00703 Modified
20	63	43.2	772	ADP87577	Adp87577 Human NOV
21	61	41.8	646	ADS20251	Ads20251 Human agg
22	61	41.1	858	ADS20230	Ads20230 Human agg
23	60	41.1	454	ABAB61917	Abab61917 Drosophila
24	59	40.4	117	AAU97558	Aau97558 Synthetic
25	59	40.4	117	AAU97553	Aau97553 Synthetic

26	59	40.4	117	5	AAU97557	Aau97557 Synthetic
27	59	40.4	117	5	AAU97555	Aau97555 Synthetic
28	59	40.4	117	5	AAU97556	Aau97556 Synthetic
29	59	40.4	117	5	AAU97552	Aau97552 Synthetic
30	59	40.4	117	5	AAU97560	Aau97560 Synthetic
31	59	40.4	118	5	AAU97554	Aau97554 Synthetic
32	59	40.4	236	8	AD019051	Ado19051 Murine an
33	59	40.4	236	8	AD019053	Ado19053 Murine an
34	58	39.7	11	6	AAE38373	Aae38373 Bp10pe t
35	58	39.7	11	8	ADS20228	Ads20228 Strept tag
36	58	39.7	19	6	ABG74881	Abg74881 Bacteriop
37	58	39.7	19	8	ABG74880	Abg74880 Bacteriop
38	58	39.7	19	8	AD139155	Adi39155 Construct
39	58	39.7	19	8	AD139156	Adi39156 Construct
40	58	39.7	31	8	ADN11208	Adn11208 Peptide m
41	58	39.7	31	8	ADN11225	Adn11225 Peptide m
42	58	39.7	31	8	ADN11207	Adn11207 Peptide m
43	58	39.7	31	8	ADN11226	Adn11226 Peptide m
44	58	39.7	109	3	AAV92150	Aav92150 C19-jun f
45	58	39.7	209	6	ADA00709	Ada00709 Modified

## ALIGNMENTS

RESULT 1  
ABP60370  
ID ABP60370 standard; peptide; 36 AA.

XX ABP60370;

DT 28-MAR-2003 (first entry)

XX Streptavidin binding peptide SEQ ID NO 11.

XX Streptavidin; protein chip; microtitre plate; detection.

XX Synthetic.

OS Key Location/Qualifiers

FT Region 9..28 /note="GGGS repeats 2-5 optionally absent, residues 13-28"

FT Region 9..12 /label="GGGS\_repeat"

FT DE10113776-A1.

XX PD 02-OCT-2002.

XX PF 21-MAR-2001; 2001DE-01013776.

XX PR 21-MAR-2001; 2001DE-01013776.

XX PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

XX PI Schmidt T;

XX DR WPI; 2003-031166/03.

XX PT New isolated peptide, useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.

XX PS Claim 8; Page 16; 18pp; German.

XX CC The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Gln, Asn or Met. (I)', which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides

CC CC



XX Sequence 396 AA;  
SQ

Query Match 57.5%; Score 84; DB 4; Length 396;  
Best Local Similarity 88.9%; Pred. No. 0.097;  
Matches 16; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 WSHPOFEK--GGSGSGGS 16  
DB 198 WSHPOFEKXAGSGSGGS 215

## RESULT 4

AAAB46423  
ID AAB46423 standard; protein; 396 AA.

AC AAB46423;

DT 06-APR-2001 (first entry)

DE Bilin binding-protein associated protein #4.

KW Bilin-binding protein; muten; BBP; digoxigenin.

OS Unidentified.

PN WO200075308-A1.

PD 14-DEC-2000.

PF 08-JUN-2000; 2000WO-DE001873.

PR 08-JUN-1999; 99DE-01026068.

PA (SKER/) SKERRA A.

PI Skerra A, Schlehuber S;

DR WPI; 2001-071071/08.

DR N-PSDB; AAF25708.

PT New muten of bilin-binding protein, useful for detecting digoxigenin  
PT being used as label in e.g. binding assays, are very selective for  
PT digoxigenin.

PS Example 2; Page 62-64; 80pp; German.

XX This invention describes novel polypeptides (I) that are mutens of bilin  
CC -binding protein (BBP), which can bind digoxigenin (Dig) or its  
CC conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and  
CC have an amino acid (aa) substitution at at least one of the positions 28,  
CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its  
CC fusion proteins, are used to bind, detect, determine, immobilize or  
CC separate Dig or its conjugates with proteins, nucleic acids,  
CC carbohydrates, other biological or synthetic macromolecules or low  
CC molecular weight compounds, particularly in assays where Dig is being  
CC used as a label. Compared with Dig-specific antibodies, (I) have a  
CC simpler structure and are easier to prepare. They have very high  
CC specificity for Dig, relative to other steroids, and fusion partners may  
CC be attached to either end without compromising their ability to bind  
CC ligand

XX Sequence 396 AA;

Query Match 57.5%; Score 84; DB 4; Length 396;  
Best Local Similarity 88.9%; Pred. No. 0.097;  
Matches 16; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 WSHPOFEK--GGSGSGGS 16  
DB 198 WSHPOFEKXAGSGSGGS 215

## RESULT 5

ADA27291  
ID ADA27291 standard; protein; 400 AA.

AC ADA27291;

DT 20-NOV-2003 (first entry)

DE Plasmid pHNGAL5 fragment protein.

KW Muten; human; neutrophil gelatinase-associated lipocalin; hNGAL; rat;  
KW alpha2-microglobulin-related protein; A2m; mouse; 24p3; uterocalin.

OS Synthetic.

FN Key Location/Qualifiers

FT Peptide 1..21 /note= "OmpA signal sequence"

FT Protein 22..400 /note= "Mature protein; fusion protein of hNGAL, Strep-  
Tag II and a fragment of phage coat protein pIII"

FT Protein 22..199 /note= "Mature modified hNGAL"

FT Peptide 200..209 /note= "Strep-Tag II affinity tag"

FT Misc-difference 210 /note= "Encoded by TAG"

FT Protein 211..400 /note= "Amino acids 217-406 of coat protein pIII"

PN WO2003029462-A1.

PD 10-APR-2003.

PF 27-SEP-2001; 2001WO-EP011213.

PR 27-SEP-2001; 2001WO-EP011213.

PA (PIER-) PIERIS PROTEOLAB AG.

PI Skerra A, Schlehuber S;

DR WPI; 2003-381639/36.

DR N-PSDB; ADA27285.

PT Generating a muten of a protein for validating the protein as drug  
PT target by subjecting the protein to mutagenesis at sequence positions  
PT corresponding to sequence positions of the human neutrophil gelatinase-  
PT associated lipocalin.

PS Disclosure; Page 57-58; 68pp; English.

XX The present invention relates to a method for generating mutens of human  
CC neutrophil gelatinase-associated lipocalin (hNGAL), rat alpha2-  
CC microglobulin-related protein (A2m) or mouse 24p3/uterocalin, by  
CC subjecting the protein to mutagenesis. In the method of generating a  
CC muten, a nucleic acid coding for the muten of the protein resulting  
CC from mutagenesis is operably fused at the 3' end with a gene coding for  
CC the coat protein pIII of a filamentous bacteriophage of the M13-family or  
CC for a fragment of the coat protein. The present sequence is the protein  
CC encoded by a fragment of plasmid pHNGAL5, used to illustrate the  
CC invention. This sequence comprises human hNGAL, a Strep-Tag II and a  
CC fragment of phage coat protein pIII.

XX Sequence 400 AA;

Query Match 57.5%; Score 84; DB 6; Length 400;  
Best Local Similarity 88.9%; Pred. No. 0.098;  
Matches 16; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 WSHPOFEK--GGSGSGGS 16  
DB 202 WSHPOFEKXAGSGSGGS 219

```

RESULT 6
ID ADA27292-standard; protein; 400 AA.
XX
AC ADA27292;
XX
DT 20-NOV-2003 (first entry)
XX
DE Plasmid pHNGAL3 fragment protein.
XX
KM Mutcin; human neutrophil gelatinase-associated lipocalin; hNGAL; rat;
XX alpha2-microglobulin-related protein; A2m; mouse; 24p3; uterocalin.
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Peptide 1..21
FT /note="Signal peptide"
FT Protein 22..400
FT /note="Mature protein; fusion protein of hNGAL, Strep-
FT tag II and a fragment of phage coat protein pIII"
FT Peptide 22..199
FT /note="Mature hNGAL"
FT /note="Strep-Tag II affinity tag"
FT Misc-difference 210
FT /note="Encoded by TAG"
FT Protein 211..400
FT /note="Amino acids 217-406 of coat protein pIII"
XX
PN WO2003029462-A1.
XX
PD 10-APR-2003.
XX
PF 27-SEP-2001; 2001WO-EP011213.
XX
PR 27-SEP-2001; 2001WO-EP011213.
XX
PA (PIER-) PIERIS PROTEOLAB AG.
XX
PI Skerra A, Schlehuber S;
XX
DR WPI; 2003-381639/36.
XX N-PSDB; ADA27286.
XX
PT Generating a mutcin of a protein for validating the protein as drug
PT target by subjecting the protein to mutagenesis at sequence positions
PT corresponding to sequence positions of the human neutrophil gelatinase-
PT associated lipocalin.
XX
PS Disclosure; Page 59-61; 68pp; English.
XX
XX The present invention relates to a method for generating mutcins of human
XX neutrophil gelatinase-associated lipocalin (hNGAL), rat alpha2-
XX microglobulin-related protein (A2m) or mouse 24p3/uterocalin, by
XX subjecting the protein to mutagenesis. In the method of generating a
XX mutcin, a nucleic acid coding for the mutcins of the protein resulting
XX from mutagenesis is operably fused at the 3' end with a gene coding for
XX the coat protein pIII of a filamentous bacteriophage of the M13-family or
XX for a fragment of the coat protein. The present sequence is the protein
XX encoded by a fragment of plasmid pHNGAL3, used to illustrate the
XX invention. This sequence comprises human hNGAL, a Strep-tag II and a
XX fragment of phage coat protein pIII.
XX
SQ Sequence 400 AA;
Query Match 57.5%; Score 84; DB 6; Length 400;
Best Local Similarity 88.9%; Pred. No. 0.098;
Matches 16; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
1 WSHPQFEK--GGGSGGGS 16

```

```

DB
|||||
202 WSHPQFEKAGGSGGGS 219
RESULT 7
ID ADA00700-standard; protein; 400 AA.
XX
AC ADA00700;
XX
DT 06-NOV-2003 (first entry).
XX
DE Modified hNGAL related fusion protein SEQ ID NO:13.
XX
KM mutcin; human neutrophil gelatinase-associated lipocalin; hNGAL;
XX rat alpha2-microglobulin-related protein; A2m; mouse 24p3/uterocalin;
XX 24p3; mutagenesis; fusion protein; cytostatic; gene therapy;
XX tumour imaging; cancer therapy.
OS Synthetic.
XX
FH Homo sapiens.
OS
FH Key Location/Qualifiers
FH Peptide 1..21
FH /label= signal
FH Protein 22..400
FH /note="modified hNGAL, Strep-tag II and phage coat
FH protein pIII fragment fusion protein"
FH Region 22..199
FH /note="mature hNGAL"
FH Region 200..209
FH /note="Strep-tag II affinity tag"
FH Misc-difference 210
FH /note="amber stop codon"
FH Region 211..400
FH /note="coat protein pIII fragment 217-406"
XX
PN WO2003029463-A2.
XX
PD 10-APR-2003.
XX
PF 18-SEP-2002; 2002WO-EP010490.
XX
PR 27-SEP-2001; 2001WO-EP011213.
XX 16-APR-2002; 2002WO-EP004223.
XX
PA (PIER-) PIERIS PROTEOLAB AG.
XX
PI Skerra A, Schlehuber S;
XX
DR WPI; 2003-372000/35.
XX N-PSDB; ADA00731.
XX
PT Generating a mutcin of a protein, e.g. human neutrophil gelatinase-
PT associated lipocalin, rat alpha2-microglobulin-related protein or mouse
PT 24p3/uterocalin, comprises subjecting the protein to mutagenesis.
XX
PS Example 3; Page 94-95; 122pp; English.
XX
XX The present invention describes a method for generating a mutcin of a
XX protein selected from a human neutrophil gelatinase-associated lipocalin
XX (hNGAL), rat alpha2-microglobulin-related protein (A2m) and a mouse
XX 24p3/uterocalin (24p3), where the mutcin has a detectable affinity to a
XX given target, comprising subjecting the protein to mutagenesis at
XX positions 33-54, 66-83, 94-106 and/or 123-136 of hNGAL, resulting in
XX mutcins of the protein. Also described: (1) a mutcin of hNGAL, A2m or
XX 24p3 having detectable binding affinity to a given target, obtained by
XX the method described above; (2) a fusion protein comprising the mutcin of
XX hNGAL, A2m or 24p3, where an enzyme, a protein or a protein domain, a
XX peptide, a signal sequence and/or an affinity tag is operably fused to
XX the amino or carboxy terminus of the mutcin; (3) a nucleic acid molecule
XX comprising a sequence encoding the mutcin of hNGAL, A2m or 24p3 or the
XX fusion protein of (2); and (4) a pharmaceutical composition comprising

```

CC the mutein of hNGAL, A2m or 24p3 or the fusion protein described above,  
 CC and a carrier. The muteins have cytostatic activity, and can be used in  
 CC gene therapy. The method is useful in generating or producing a mutein of  
 CC hNGAL, A2m or 24p3 or a fusion protein. The mutein of hNGAL, A2m or 24p3  
 CC or the fusion protein is useful in detecting a given target by contacting  
 CC the mutein with a sample suspected of containing the given target under  
 CC conditions that allow complex formation between the mutein and the given  
 CC target, and determining the complexed mutein by a suitable signal. The  
 CC given target is a protein or protein domain, a peptide, a nucleic acid  
 CC molecule, an organic molecule or a metal complex and the detection is  
 CC carried out for validation of the protein as a pharmacological drug  
 CC target. The mutein may also be used in medicine, e.g. for tumour imaging  
 CC or directly for cancer therapy. The present sequence represents a  
 CC modified hNGAL, Strep-tag II and phage coat protein pIII fragment fusion  
 CC protein given in the exemplification of the present invention.

XX SQ Sequence 400 AA;

Query Match 57.5%; Score 84; DB 6; Length 400;  
 Best Local Similarity 88.9%; Pred. No. 0.098;  
 Matches 16; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 WSHPOFEK--GGSGSGGS 16  
 |||||  
 DB 202 WSHPOFEKQAGGSGSGGS 219

RESULT 8  
 ADA00707  
 ID ADA00707 standard; protein; 400 AA.

XX AC ADA00707;

XX DT 06-NOV-2003 (first entry)

XX DE Modified hNGAL related fusion protein SEQ ID NO:20.

KM mutein; human neutrophil gelatinase-associated lipocalin; hNGAL;  
 KM rat alpha2-microglobulin-related protein; A2m; mouse 24p3/uterocalin;  
 KM 24p3; mutagenesis; fusion protein; cytostatic; gene therapy;  
 KM tumour imaging; cancer therapy.

OS Synthetic.  
 OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..21  
 FT /label= signal

FT Protein 22..400  
 FT /note= "modified hNGAL, Strep-tag II and phage coat  
 protein pIII fragment fusion protein"

FT Region 22..199  
 FT /note= "mature hNGAL"

FT Region 200..209  
 FT /note= "Strep-tag II affinity tag"

FT Misc-difference 210  
 FT /note= "amber stop codon"

FT Region 211..400  
 FT /note= "coat protein pIII fragment 217-406"

XX PN WO2003029463-A2.

XX PD 10-APR-2003.

XX PF 18-SEP-2002; 2002WO-EP010490.

XX PR 27-SEP-2001; 2001WO-EP011213.

XX PR 16-APR-2002; 2002WO-EP004223.

XX (PIER-) PIERIS PROTEOLAB AG.

XX Skerra A, Schlehuber S;

XX PI

XX FT

DR WPI; 2003-372000/35.  
 DR N-PSDB; ADA00706.

PT Generating a mutein of a protein, e.g. human neutrophil gelatinase-  
 PT associated lipocalin, rat alpha2-microglobulin-related protein or mouse  
 PT 24p3/uterocalin, comprises subjecting the protein to mutagenesis.

PS Disclosure; Page 104-105; 122pp; English.

XX The present invention describes a method for generating a mutein of a  
 CC protein selected from a human neutrophil gelatinase-associated lipocalin  
 CC (hNGAL), rat alpha2-microglobulin-related protein (A2m) and a mouse  
 CC 24p3/uterocalin (24p3), where the mutein has a detectable affinity to a  
 CC given target, comprising subjecting the protein to mutagenesis at  
 CC positions 33-34, 66-83, 94-106 and/or 123-136 of hNGAL, resulting in  
 CC muteins of the protein. Also described: (1) a mutein of hNGAL, A2m or  
 CC 24p3 having detectable binding affinity to a given target, obtained by  
 CC the method described above; (2) a fusion protein comprising the mutein of  
 CC hNGAL, A2m or 24p3, where an enzyme, a protein or a protein domain, a  
 CC peptide, a signal sequence and/or an affinity tag is operably fused to  
 CC the amino or carboxy terminus of the mutein; (3) a nucleic acid molecule  
 CC comprising a sequence encoding the mutein of hNGAL, A2m or 24p3 or the  
 CC fusion protein of (2); and (4) a pharmaceutical composition comprising  
 CC the mutein of hNGAL, A2m or 24p3 or the fusion protein described above,  
 CC and a carrier. The muteins have cytostatic activity, and can be used in  
 CC gene therapy. The method is useful in generating or producing a mutein of  
 CC hNGAL, A2m or 24p3 or a fusion protein. The mutein of hNGAL, A2m or 24p3  
 CC or the fusion protein is useful in detecting a given target by contacting  
 CC the mutein with a sample suspected of containing the given target under  
 CC conditions that allow complex formation between the mutein and the given  
 CC target, and determining the complexed mutein by a suitable signal. The  
 CC given target is a protein or protein domain, a peptide, a nucleic acid  
 CC molecule, an organic molecule or a metal complex and the detection is  
 CC carried out for validation of the protein as a pharmacological drug  
 CC target. The mutein may also be used in medicine, e.g. for tumour imaging  
 CC or directly for cancer therapy. The present sequence represents a  
 CC modified hNGAL, Strep-tag II and phage coat protein pIII fragment fusion  
 CC protein given in the exemplification of the present invention.

XX SQ Sequence 400 AA;

Query Match 57.5%; Score 84; DB 6; Length 400;  
 Best Local Similarity 88.9%; Pred. No. 0.098;  
 Matches 16; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 WSHPOFEK--GGSGSGGS 16  
 |||||  
 DB 202 WSHPOFEKQAGGSGSGGS 219

RESULT 9  
 ADA00701  
 ID ADA00701 standard; protein; 400 AA.

XX AC ADA00701;

XX DT 06-NOV-2003 (first entry)

XX DE Modified hNGAL related fusion protein SEQ ID NO:14.

KM mutein; human neutrophil gelatinase-associated lipocalin; hNGAL;  
 KM rat alpha2-microglobulin-related protein; A2m; mouse 24p3/uterocalin;  
 KM 24p3; mutagenesis; fusion protein; cytostatic; gene therapy;  
 KM tumour imaging; cancer therapy.

OS Synthetic.  
 OS Homo sapiens.

XX FH Key Location/Qualifiers

FT Peptide 1..21  
 FT /label= signal

FT Protein 22..400  
 FT /note= "modified hNGAL, Strep-tag II and phage coat



XX	Homo sapiens.
OS	Unidentified.
XX	
Key	Location/Qualifiers
FH	Misc-difference 475..485
FT	/note="Residues corresponding to positions 687-837 in the wild-type replaced by Strep tag"
FN	
MO	2004011637-A2.
XX	
PD	05 - FEB - 2004.
XX	
PF	29 - JUL - 2003; 2003WO-US023464.
PR	
ER	29 - JUL - 2002; 2002US-0398721P.
XX	
(AMHP ) WYETH.	
PA	(CORC/) CORCORAN C J.
PA	(PLANN/) PLANNERY C R.
PA	(ZENG/) ZENG W.
PA	(RACI/) RACIE L A.
PA	(MCDO/) MCDONAGH T.
PA	(FREEB/) FREEMAN B A.
PA	(GEOR/) GEORGIADIS K E.
PA	(LAVN/) LAVALLIE E R.
XA	
P1	Corcoran CJ, Plannery CR, Zeng W, Racie LA, McDonagh T;
P1	Freeman BA, Georgiadis KE, Lavallie ER,
DR	WPI; 2004-143860/14.
XX	
PT	New isolated, modified ADAMTS4 (aggrecanase) protein with improved stability useful for identifying inhibitors of the enzyme activity for treating aggrecanase-associated conditions, including osteoarthritis.
PS	Claim 9; SEQ ID NO 47; 117pp; English.
CC	The invention relates to a novel isolated, modified ADAMTS4 (a chondrointegrin-like and metalloproteinase (reprolysin type) with thrombospondin type I motif 4) protein with improved stability compared to a naturally occurring, full-length ADAMTS4 protein, where the modified protein differs from the naturally-occurring, full-length ADAMTS4 protein by at least one amino acid. ADAMTS proteins are a subfamily of zinc metalloproteinases and include aggrecanases amongst their members. The protein of the invention demonstrates osteopathic, anti-inflammatory, antiarthritic, antirheumatic and cytostatic activities and may be useful for treating aggrecanase-associated conditions, including osteoarthritis, glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic arthritis, periodontal disease and Crohn's disease. The current sequence is that of the furin-processed human aggrecanase ADAMTS4 truncated protein with Strep tag of the invention.
SQ	Sequence 485 AA;
Oy	Query Match 49.0%; Score 71.5; DB 8; Length 485; Best Local Similarity 46.7%; Pred. No. 3.3; Matches 14; Conservative 3; Mismatches 6; Indels 7; Gaps 1
Dd	2 SHPOFEK-----GGSGCGSMWSHPOFEK 24     :   :   :   :   :   :   :   :   456 SKKKPDFCKMVCGGDSGCSGSASMSHPOFEK 485
ID	ADSS20227 standard; protein; 697 AA.
XX	

AC		Abs20227;	
XX			
DT		18-NOV-2004	(first entry)
DE		Human aggrecanase ADAMTS4 truncated protein with Streptag - SEQ ID NO 24.	
KW		ADAMTS4; a disintegrin-like and metalloprotease; thrombospondin type I motif 4; reprolysin; zinc metalloprotease; aggrecanase; osteopontic; anti-inflammatory; antirheumatic; antihemostatic; osteoarthritis; glioma; cancer; inflammatory joint; rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease; human; enzyme; chromosome 1q21-q23; truncation; Streptag.	
XX		Homo sapiens.	
OS		Undetermined.	
FH	Key	Location/Qualifiers	
FT	Misc-difference	687..697 /note= "Wild-type residues 687-837 replaced by Streptag"	
XX			
EN		WO2004011637-A2.	
PD		05-FEB-2004.	
XX			
PF		29-JUL-2003; 2003WO-US023484.	
PR		29-JUL-2002; 2002US-0398721P.	
PA	(AHNP ) WYETH.		
PA	(CORC/) CORCORAN C J.		
PA	(FLANN) FLANNERY C R.		
PA	(ZENG/) ZENG W.		
PA	(RACI/) RACIE L A.		
PA	(MCDO/) MCDONAGH T.		
PA	(FREEB/) FREEMAN B A.		
PA	(GROGI/) GEORGIADIS K E.		
PA	(LAVALL) LAVALLIE E R.		
XI			
PI	Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T,		
PI	Freeman BA, Georgiadis KE, Lavallie ER;		
DR			
XX		WPt, 2004-143860/14.	
PT	New isolated, modified ADAMTS4 (aggrecanase) protein with improved stability useful for identifying inhibitors of the enzyme activity for treating aggrecanase-associated conditions, including osteoarthritis.		
PS	Claim 9; SEQ ID NO 24; 117pp; English.		
CC	The invention relates to a novel isolated, modified ADAMTS4 (a disintegrin-like and metalloprotease (reprolysin type)) with thrombospondin type I motif 4) protein with improved stability compared to a naturally occurring, full-length ADAMTS4 protein, where the modified protein differs from the naturally-occurring, full-length ADAMTS4 protein by at least one amino acid. ADAMTS proteins are a subfamily of zinc metalloproteinases and include aggrecanases amongst their members. The protein of the invention demonstrates osteopathic, anti-inflammatory, antiarthritic, antihemostatic and cytoskeletal activities and may be useful for treating aggrecanase-associated conditions, including osteoarthritis, glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic arthritis, periodontal disease and Crohn's disease. The current sequence is that of the human aggrecanase ADAMTS4 truncated protein with Streptag of the invention.		
SQ	Sequence 697 AA:		
Query Match	49.0%; Score 71.5; DB 8; Length 697;		
Best Local Similarity	46.7%; Pred.No. 4.8;		
Matches	14; Conservative 3; Mismatches 6; Indels 7; Gaps 1.		
QY	2 SHPOPEK-----CGSGSCGSWSHPPOEK 24		
DB	:: ::  ::      668 SKKKFDKMVCGGDSGCCSASWSHPPOEK 697		

RESULT 13  
 ADS20243  
 ID ADS20243 standard; protein; 845 AA.  
 XX  
 AC ADS20243;  
 XX  
 DT 18-NOV-2004 (first entry)  
 XX  
 DE Human aggrecanase ADAMTS4 mutant E362Q protein with inserted Strep tag.  
 XX  
 KW ADAMTS4, a disintegrin-like and metalloprotease;  
 XX  
 KW thrombospondin type 1 motif 4; reprotysin; zinc metalloprotease;  
 KW aggrecanase; osteopontin; antiinflammatory; antiarthritic; antineumatic;  
 KW cytosolic; osteoarthritis; glioma; cancer; inflammatory joint;  
 KW rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;  
 KW human; enzyme; chromosome 1q21-q23; Strep tag; mutant; mutein.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 OS Unidentified.  
 XX  
 FT Key Location/Qualifiers  
 FT Misc-difference 362 /note="Wild-type Glu substituted by Gln"  
 FT Misc-difference 520..527 /note="Strep tag peptide 2 inserted"  
 FT  
 PN WO2004011637-A2.  
 XX  
 PD 05-FEB-2004.  
 XX  
 PF 29-JUL-2003; 2003WO-US023484.  
 XX  
 PR 29-JUL-2002; 2002US-0398721P.  
 XX  
 PA (AMHP) WYETH.  
 PA (CORC) CORCORAN C J.  
 PA (FLAN) FLANNERY C R.  
 PA (ZENG) ZENG W.  
 PA (RACI) RACIE L A.  
 PA (MCDON) MCDONAGH T.  
 PA (FERRE) FREEMAN B A.  
 PA (GEOR) GEORGIADIS K E.  
 PA (LAVA) LAVALLIE E R.  
 XX  
 PI Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T;  
 PI Freeman BA, Georgiadis KE, Lavallie ER;  
 XX  
 DR WPI, 2004-143860/14.  
 XX  
 PT New isolated, modified ADAMTS4 (aggrecanase) protein with improved  
 PT stability useful for identifying inhibitors of the enzyme activity for  
 PT treating aggrecanase-associated conditions, including osteoarthritis.  
 XX  
 PS Claim 9; SEQ ID NO 40; 117pp; English.  
 XX  
 CC The invention relates to a novel isolated, modified ADAMTS4 (a  
 CC disintegrin-like and metalloprotease (reprotysin type) with  
 CC thrombospondin type 1 motif 4) protein with improved stability compared  
 CC to a naturally occurring, full-length ADAMTS4 protein, where the modified  
 CC protein differs from the naturally-occurring, full-length ADAMTS4 protein  
 CC by at least one amino acid. ADAMTS proteins are a subfamily of zinc  
 CC metalloproteases and include aggrecanases amongst their members. The  
 CC protein of the invention demonstrates osteopontin, antiinflammatory,  
 CC antiarthritic, antineumatic and cytostatic activities and may be useful  
 CC for treating aggrecanase-associated conditions, including osteoarthritis,  
 CC glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic  
 CC arthritis, periodontal disease and Crohn's disease. The current sequence  
 CC is that of the human aggrecanase ADAMTS4 mutant E362Q protein with  
 CC inserted Strep tag of the invention.  
 XX

SEQ Sequence 845 AA;  
 Query Match 47.6%; Score 69.5; DB 8; Length 845;  
 Best Local Similarity 72.2%; Pred. No. 9.8;  
 Matches 13; Conservative 0; Mismatches 4; Indels 1; Gaps 1;  
 QY 1 WSHPOFEKGGSG-GGSW 17  
 |||||  
 DB 520 WSHPOFEKAGGWGWPW 537  
 RESULT 14  
 ID ABP60362  
 XX ABP60362 standard; peptide; 24 AA.  
 XX  
 AC ABP60362;  
 XX  
 DT 28-MAR-2003 (first entry)  
 XX  
 DE Streptavidin binding peptide SEQ ID NO 3.  
 XX  
 KW Streptavidin; protein chip; microtitre plate; detection.  
 XX  
 OS Synthetic.  
 OS  
 FT Key Location/Qualifiers  
 FT Misc-difference 9..16 /label= unknown  
 FT  
 PN DE10113776-A1.  
 XX  
 PD 02-OCT-2002.  
 XX  
 PF 21-MAR-2001; 2001DE-01013776.  
 XX  
 PR 21-MAR-2001; 2001DE-01013776.  
 XX  
 PA (BIOA) INST BIOANALYTIC GMBH GOETTINGEN.  
 XX  
 PI Schmidt T;  
 XX  
 DR WPI, 2003-031166/03.  
 XX  
 PT New isolated peptide, useful as affinity purification tag for recombinant  
 PT protein, comprises at least two high-affinity streptavidin-binding  
 PT modules.  
 XX  
 PS Disclosure; Page 4; 18pp; German.  
 XX  
 CC The invention relates to an isolated peptide (I) comprising at least two  
 CC individual modules separated by 0-50 amino acids, with each containing at  
 CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain  
 CC streptavidin binding modules, are useful as affinity handles for  
 CC purification of recombinant fusion proteins (FP), also for detecting FP,  
 CC e.g. on protein chips or microtitre plates. The modules in (I) bind  
 CC strongly to streptavidin, with a co-operative effect that provides  
 CC stronger binding than a single tag, but are displaced by a competitor.  
 CC (I) does not interfere with the function of attached proteins (II) (so it  
 CC may not be essential to remove it); facilitates detection and has easily  
 CC controllable binding properties. (I) is particularly used for purifying  
 CC FP from dilute solution in batch formats (which use simpler apparatus  
 CC than column methods and result in lower loss of FP). The present sequence  
 CC is that of a streptavidin binding peptide disclosed with the invention  
 XX  
 SQ Sequence 24 AA;  
 Query Match 47.3%; Score 69; DB 6; Length 24;  
 Best Local Similarity 46.9%; Pred. No. 0.33;  
 Matches 15; Conservative 0; Mismatches 1; Indels 16; Gaps 2;  
 QY 1 WSHPOFEKGGSGGGS-----WSHPOFEK 24  
 |||||  
 DB 1 WSHPOFE-----KXXXXXXXXXWSHPOFEK 24



## RESULT 15

AAO19986 standard; protein; 42 AA.

AAO19986;

24-MAY-2002 (first entry)

Protein of (double hexahistidine) - flexible linker upstream.

Protein array; cell free system; gridded format; covalent; non-covalent; ligand; phage display library; cellular protein expression profile; ribosome display library; post-translational modification; hexahistidine; flexible linker.

Unidentified.

WO200214860-A1.

21-FEB-2002.

15-AUG-2001; 2001WO-GB003657.

15-AUG-2000; 2000GB-00020016.

01-DEC-2000; 2000GB-00029309.

16-MAR-2001; 2001GB-00006610.

07-JUN-2001; 2001GB-00013883.

14-JUL-2001; 2001GB-00017232.

(DISC-) DISCERNA LTD.

He M, Tauesig MJ;

WPI; 2002-241989/29.

N-PSDB; AAK98912.

Protein arrays useful for identifying interactions of arrayed proteins

with other molecules, in which proteins are produced by in vitro

synthesis using cell-free transcription and translation systems.

Disclosure; Fig 3B; 61pp; English.

The invention relates to a method for producing a protein array made by

transcription and translation of DNA, or by translation of mRNA, by a

cell free system in vitro to produce individual proteins, domains or

peptides distributed in a gridded format. The peptides of the invention

contain amino acid sequences enabling covalent/non-covalent attachment to

a surface/bead, such that the peptides can be arranged in the grid format

after interaction with suitable ligands or reagents on the surface. The

method of the invention is useful for identifying interactions of the

arrayed proteins with other molecules, including antibodies, other

proteins or domains, peptides, small ligands, cell extracts and nucleic

acids. The protein array is also useful for identifying interactions of

arrayed proteins with other molecules displayed in a library, such as a

phage display library or ribosome display library. The protein array is

useful for studying cellular protein expression profiles and post-

translational modification of cellular proteins. This sequence of the

invention represents a protein of a double hexahistidine with a flexible

linker sequence upstream

Sequence 42 AA;

Query Match 45.2%; Score 66; DB 5; Length 42;

Best Local Similarity 59.1%; Pred. No. 1.3;

Matches 13; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

QY 1 WSHPOF-----EKGGSGGGGS 16

DB 10 WRNPOFGGHHHHHGGSGGGGS 31

Search completed: March 2, 2005, 13:02:48  
 Job time : 109.098 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 74.6341 Seconds  
(without alignments)  
105.489 Million cell updates/sec

Title: SEQ11  
Perfect score: 146  
Sequence: 1 wnhpfekgsgsgsggwhpfek 24

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	117	80.1	36	US-10-026-578B-11	Sequence 11, Appl
2	71.5	49.0	485	US-10-628-432-47	Sequence 47, Appl
3	71.5	49.0	697	US-10-628-432-24	Sequence 24, Appl
4	69.5	47.6	845	US-10-628-432-40	Sequence 40, Appl
5	69	45.3	24	US-10-026-578B-3	Sequence 3, Appl
6	66	45.2	42	US-10-344-607-16	Sequence 16, Appl
7	63	43.2	36	US-10-026-578B-10	Sequence 10, Appl
8	63	43.2	42	US-10-344-607-20	Sequence 20, Appl
9	61	41.8	646	US-10-628-432-49	Sequence 49, Appl
10	61	41.8	858	US-10-628-432-27	Sequence 27, Appl
11	61	41.8	929	US-10-437-963-167129	Sequence 4, Appl
12	59	40.4	117	US-09-977-137A-4	Sequence 4, Appl
13	59	40.4	117	US-09-977-137A-5	Sequence 5, Appl

14	59	40.4	117	10	US-09-977-137A-7	Sequence 7, Appl
15	59	40.4	117	10	US-09-977-137A-8	Sequence 8, Appl
16	59	40.4	117	10	US-09-977-137A-9	Sequence 9, Appl
17	59	40.4	117	10	US-09-977-137A-10	Sequence 10, Appl
18	59	40.4	117	10	US-09-977-137A-11	Sequence 11, Appl
19	59	40.4	118	10	US-09-977-137A-12	Sequence 12, Appl
20	58.5	40.1	81	16	US-10-437-963-197510	Sequence 6, Appl
21	58	39.7	11	15	US-10-354-963-29	Sequence 29, Appl
22	58	39.7	11	16	US-10-628-432-25	Sequence 25, Appl
23	58	39.7	109	15	US-10-302-100B-41	Sequence 41, Appl
24	58	39.7	661	15	US-10-354-963-8	Sequence 8, Appl
25	58	39.7	763	15	US-10-358-283-15	Sequence 15, Appl
26	57.5	39.4	633	16	US-10-628-432-53	Sequence 53, Appl
27	57	39.0	10	15	US-10-147-211A-20	Sequence 20, Appl
28	56.5	38.7	253	17	US-10-728-246-13	Sequence 13, Appl
29	56	38.4	21	9	US-09-809-517A-30	Sequence 30, Appl
30	56	38.4	22	9	US-09-809-517A-33	Sequence 33, Appl
31	56	38.4	22	9	US-09-809-517A-31	Sequence 31, Appl
32	56	38.4	25	9	US-09-809-517A-34	Sequence 34, Appl
33	56	38.4	129	16	US-10-437-963-161505	Sequence 161505,
34	56	38.4	245	17	US-10-887-228A-1	Sequence 1, Appl
35	56	38.4	246	17	US-10-887-228A-9	Sequence 9, Appl
36	56	38.4	252	17	US-10-887-228A-5	Sequence 5, Appl
37	56	38.4	259	14	US-10-304-630-33	Sequence 33, Appl
38	56	38.4	582	16	US-10-437-963-193053	Sequence 193053,
39	56	38.4	795	16	US-10-437-963-108411	Sequence 108411,
40	55.5	38.0	286	16	US-10-437-963-193287	Sequence 193287,
41	55.5	38.0	592	16	US-10-322-281-559	Sequence 559, App
42	55	37.7	204	16	US-10-437-963-161117	Sequence 161117,
43	55	37.7	247	16	US-10-437-963-148362	Sequence 148362,
44	55	37.7	479	16	US-10-437-963-151628	Sequence 151628,
45	55	37.7	959	16	US-10-437-963-181639	Sequence 181639,

## ALIGNMENTS

RESULT 1  
US-10-026-578B-11  
Sequence 11, Application US/10026578B  
GENERAL INFORMATION:  
APPLICANT: IBA (GmbH)  
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta  
FILE REFERENCE: 100810.01US1  
CURRENT APPLICATION NUMBER: US/10/026,578B  
CURRENT FILING DATE: 2002-11-11  
PRIOR APPLICATION NUMBER: DE 101 13 776.1  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: PCT/EP01/11846  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: Patentin version 3.1  
SEQ ID NO 11  
LENGTH: 36  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Synthetic Peptide  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (9)..(28)  
OTHER INFORMATION: Of the five (GGGS) repeats between position 9 and 28, one or more  
OTHER INFORMATION: repeats may be missing. However, at least one (GGGS) repeat will  
FEATURE:  
NAME/KEY: misc feature  
OTHER INFORMATION: Artificial Sequence represents peptide binding module  
US-10-026-578B-11  
Query Match 80.1%; Score 117; DB 14; Length 36;  
Best Local Similarity 63.9%; Pred. No. 1.5e-06;



```

; NAME/KEY: MISC FEATURE
; LOCATION: (16)_(16)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-3

Query Match          47.3%; Score 69; DB 14; Length 24;
Best Local Similarity 46.9%; Pred. No. 0.31;
Matches 15; Conservative 0; Mismatches 1; Indels 16; Gaps 2;

QY      1 WSHPOFEKGGSGSGS-----WSHPQFEK 24
Db      1 WSHPOFE-----KXXXXXXXXXWSHPQFEK 24

RESULT 6
US-10-344-607-16
; Sequence 16, Application US/10344607
; Publication No. US20040161748A1
; GENERAL INFORMATION:
; APPLICANT: HE, MINGYUE
; APPLICANT: TAUSSIG, MICHAEL JOHN
; TITLE OF INVENTION: FUNCTIONAL PROTEIN ARRAYS
; FILE REFERENCE: 37945-0049
; CURRENT APPLICATION NUMBER: US/10/344,607
; CURRENT FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: PCT/GB01/03657
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 16
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-344-607-16

Query Match          45.2%; Score 66; DB 16; Length 42;
Best Local Similarity 59.1%; Pred. No. 1.1;
Matches 13; Conservative 0; Mismatches 3; Indels 6; Gaps 1;

QY      1 WSHPOF-----EKGGSGSGGS 16
Db      10 WRHPQFGHHHHHGGSGSGGS 31

RESULT 7
US-10-026-578B-10
; Sequence 10, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 10
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
```

```

; NAME/KEY: MISC FEATURE
; LOCATION: (9)_(28)
; OTHER INFORMATION: X represents a single amino acid at each of the positions indicat
; OTHER INFORMATION: ed; some of the amino acids may be missing. Where amino acids are
; OTHER INFORMATION: missing, the total numbers of x will be no less than 5
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-10

Query Match          43.2%; Score 63; DB 14; Length 36;
Best Local Similarity 34.1%; Pred. No. 2.2;
Matches 15; Conservative 0; Mismatches 1; Indels 28; Gaps 2;

QY      1 WSHPOFEKGGSGSGS-----WSHPQFEK 24
Db      1 WSHPOFE-----KXXXXXXXXXXXXXXXXXXXXWSHPQFEK 36

RESULT 8
US-10-344-607-20
; Sequence 20, Application US/10344607
; Publication No. US20040161748A1
; GENERAL INFORMATION:
; APPLICANT: HE, MINGYUE
; APPLICANT: TAUSSIG, MICHAEL JOHN
; TITLE OF INVENTION: FUNCTIONAL PROTEIN ARRAYS
; FILE REFERENCE: 37945-0049
; CURRENT APPLICATION NUMBER: US/10/344,607
; CURRENT FILING DATE: 2003-02-13
; PRIOR APPLICATION NUMBER: PCT/GB01/03657
; PRIOR FILING DATE: 2001-08-15
; NUMBER OF SEQ ID NOS: 20
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 20
; LENGTH: 42
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
US-10-344-607-20

Query Match          43.2%; Score 63; DB 16; Length 42;
Best Local Similarity 54.5%; Pred. No. 2.5;
Matches 12; Conservative 1; Mismatches 1; Indels 8; Gaps 1;

QY      9 GGGSGGG-----WSHPQF 22
Db      13 GGGSGGGHHHHHSRMRHPQF 34

RESULT 9
US-10-628-432-49
; Sequence 49, Application US/10628432
; Publication No. US20040142863A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Modified ADAMTS4 molecules
; FILE REFERENCE: AM101378
; CURRENT APPLICATION NUMBER: US/10/628,432
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49
; LENGTH: 646
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: furin-processed construct E
US-10-628-432-49
```

Query Match 41.8%; Score 61; DB 16; Length 646;  
Best Local Similarity 71.4%; Pred. No. 49;  
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 11 GSGGSGSMHPQPEK 24  
DB 633 GRKGSAMSHPOPEK 646

## RESULT 10

US-10-628-432-27  
; Sequence 27, Application US/10628432  
; Publication No. US20040142863A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; TITLE OF INVENTION: Modified ADAMTS4 molecules  
; FILE REFERENCE: AM101378  
; CURRENT APPLICATION NUMBER: US/10/628,432  
; CURRENT FILING DATE: 2003-07-29  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO: 27  
; LENGTH: 858  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: modified ADAMTS4 molecule  
US-10-628-432-27

Query Match 41.8%; Score 61; DB 16; Length 858;  
Best Local Similarity 71.4%; Pred. No. 63;  
Matches 10; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 11 GSGGSGSMHPQPEK 24  
DB 845 GRKGSAMSHPOPEK 858

## RESULT 11

US-10-437-963-167129  
; Sequence 167129, Application US/10437963  
; Publication No. US20040123343A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa, Thomas J.  
; APPLICANT: Kovallig, David K.  
; APPLICANT: Zhou, Yihua  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Wu, Wei  
; APPLICANT: Boukharov, Andrey A.  
; APPLICANT: Barbazuk, Brad  
; APPLICANT: Li, Ping  
; TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53221)B  
; CURRENT APPLICATION NUMBER: US/10/437,963  
; CURRENT FILING DATE: 2003-05-14  
; NUMBER OF SEQ ID NOS: 204966  
; SEQ ID NO: 167129  
; LENGTH: 929  
; TYPE: PRT  
; ORGANISM: Oryza sativa  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT4530\_65770C.1.pap  
US-10-437-963-167129

Query Match 41.8%; Score 61; DB 16; Length 929;  
Best Local Similarity 52.9%; Pred. No. 67;  
Matches 9; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

QY 1 WSHPOPEKGGGSGGGSW 17  
DB 537 WRSFAMDHGGGSGGGSW 553

RESULT 12  
US-09-977-137A-4  
; Sequence 4, Application US/09977137A  
; Publication No. US20030104524A1  
; GENERAL INFORMATION:  
; APPLICANT: Sumner, Anne O.  
; APPLICANT: Caguiat, Jonathan  
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and  
; TITLE OF INVENTION: Methods

; FILE REFERENCE: 79-00  
; CURRENT APPLICATION NUMBER: US/09/977,137A  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: US 60/240,465  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 4  
; LENGTH: 117  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: chelon  
US-09-977-137A-4

Query Match 40.4%; Score 59; DB 10; Length 117;  
Best Local Similarity 64.7%; Pred. No. 18;  
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 8 KGGGSGGSMHPQPEK 24  
DB 101 KGNVSCPSAMSHPOPEK 117

## RESULT 13

US-09-977-137A-5  
; Sequence 5, Application US/09977137A  
; Publication No. US20030104524A1  
; GENERAL INFORMATION:  
; APPLICANT: Sumner, Anne O.  
; APPLICANT: Caguiat, Jonathan  
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and  
; TITLE OF INVENTION: Methods  
; FILE REFERENCE: 79-00  
; CURRENT APPLICATION NUMBER: US/09/977,137A  
; CURRENT FILING DATE: 2001-10-12  
; PRIOR APPLICATION NUMBER: US 60/240,465  
; PRIOR FILING DATE: 2000-10-12  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO: 5  
; LENGTH: 117  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: chelon  
US-09-977-137A-5

Query Match 40.4%; Score 59; DB 10; Length 117;  
Best Local Similarity 64.7%; Pred. No. 18;  
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 8 KGGGSGGSMHPQPEK 24  
DB 101 KGNVSCPSAMSHPOPEK 117

## RESULT 14

US-09-977-137A-7  
; Sequence 7, Application US/09977137A  
; Publication No. US20030104524A1  
; GENERAL INFORMATION:  
; APPLICANT: Sumner, Anne O.

```

; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-7

```

```

Query Match          40.4%; Score 59; DB 10; Length 117;
Best Local Similarity 64.7%; Pred. No. 18;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy      8 KGGGCGGGSWSHPQFEK 24
      |||:|||||
Db      101 KGNVSCPSAWSHPQFEK 117

```

```

RESULT 15
US-09-977-137A-8
; Sequence 8, Application US/0977137A
; Publication No. US20030104524A1
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-8

```

```

Query Match          40.4%; Score 59; DB 10; Length 117;
Best Local Similarity 64.7%; Pred. No. 18;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

```

```

Qy      8 KGGGCGGGSWSHPQFEK 24
      |||:|||||
Db      101 KGNVSCPSAWSHPQFEK 117

```

```

Search completed: March 2, 2005, 14:18:52
Job time : 74.6341 secs

```

*This Page Blank (uspto)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using ew model

Run on: March 2, 2005, 12:20:44 ; Search time 27.3171 Seconds  
(without alignment)  
65.585 Million cell updates/sec

Title: SEQ11  
Perfect score: 146  
Sequence: 1 wnhpbfeksgsgsgshpbfek 24

Scoring table:  
Gapop 10.0, Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/6CTUS\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/Backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	59	40.4	117	4	US-09-977-137A-4
2	59	40.4	117	4	US-09-977-137A-5
3	59	40.4	117	4	US-09-977-137A-7
4	59	40.4	117	4	US-09-977-137A-8
5	59	40.4	117	4	US-09-977-137A-9
6	59	40.4	117	4	US-09-977-137A-10
7	59	40.4	117	4	US-09-977-137A-12
8	59	40.4	118	4	US-09-977-137A-6
9	58	39.7	109	4	US-09-407-687-41
10	56	38.4	21	4	US-09-809-517A-30
11	56	38.4	22	4	US-09-809-517A-33
12	56	38.4	24	4	US-09-809-517A-31
13	56	38.4	25	4	US-09-809-517A-34
14	56	38.4	259	4	US-09-431-887-33
15	56	38.4	660	2	US-08-770-761A-2
16	56	38.4	662	2	US-08-770-761A-5
17	54	37.0	705	2	US-08-770-761A-7
18	54	37.0	117	4	US-09-977-137A-11
19	53	36.3	287	4	US-09-248-796A-15613
20	53	36.3	99	4	US-09-407-687-43
21	52.5	36.0	286	4	US-09-248-796A-15692
22	52.5	36.0	450	4	US-09-849-016-7527
23	52	35.6	10	4	US-09-809-517A-6
24	52	35.6	521	4	US-09-252-991A-18266
25	51	34.9	563	4	US-09-600-991A-4
26	51	34.9	8	3	US-08-948-097-2
27	51	34.9	8	3	US-09-382-950-7

28	51	34.9	8	3	US-09-382-736B-8	Sequence 8, Appl
29	51	34.9	8	4	US-09-619-103-9	Sequence 9, Appl
30	51	34.9	8	4	US-10-104-218-5	Sequence 5, Appl
31	51	34.9	8	4	US-09-809-517A-9	Sequence 9, Appl
32	51	34.9	26	2	US-08-596-387B-97	Sequence 97, Appl
33	51	34.9	26	2	US-09-067-615-97	Sequence 97, Appl
34	51	34.9	26	5	PCR-US95-09816A-97	Sequence 97, Appl
35	51	34.9	179	4	US-09-270-767-38134	Sequence 38134, A
36	51	34.9	179	4	US-09-270-767-35351	Sequence 53351, A
37	51	34.9	219	4	US-09-809-517A-24	Sequence 24, Appl
38	51	34.9	219	4	US-09-809-517A-27	Sequence 27, Appl
39	51	34.9	229	4	US-09-248-796A-17127	Sequence 17127, A
40	51	34.9	238	4	US-09-495-880A-42	Sequence 42, Appl
41	51	34.9	254	4	US-09-270-767-60227	Sequence 60227, A
42	51	34.9	266	4	US-09-495-880A-26	Sequence 26, Appl
43	51	34.9	328	4	US-09-270-767-44768	Sequence 44768, A
44	51	34.9	699	4	US-09-252-991A-17780	Sequence 17780, A
45	50.5	34.6	256	3	US-09-128-450-22	Sequence 22, Appl

## ALIGNMENTS

```
RESULT 1
US-09-977-137A-4
; Sequence 4, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4

Query Match      40.4% Score 59; DB 4; Length 117;
Best Local Similarity 64.7% Pred. No. 3.2;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      8 KGGSGGGGSMHPQFEK 24
Db      101 KGNVSCPSMSHPQFEK 117

RESULT 2
US-09-977-137A-5
; Sequence 5, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
```

LENGTH: 117  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: chelon  
US-09-977-137A-5

Query Match 40.4%; Score 59; DB 4; Length 117;  
Best Local Similarity 64.7%; Pred. No. 3.2;  
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 8 KGGSGGSGSMWHPQFEK 24  
DB 101 KGNVSCPSAMWHPQFEK 117

RESULT 3  
US-09-977-137A-7  
Sequence 7, Application US/09977137A  
Patent No. 6750042  
GENERAL INFORMATION:  
APPLICANT: Summers, Anne O.  
APPLICANT: Caguiat, Jonathan  
TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and  
FILE REFERENCE: 79-00  
CURRENT APPLICATION NUMBER: US/09/977,137A  
PRIOR FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: US 60/240,465  
PRIOR FILING DATE: 2000-10-12  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 7  
LENGTH: 117  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: chelon  
US-09-977-137A-7

Query Match 40.4%; Score 59; DB 4; Length 117;  
Best Local Similarity 64.7%; Pred. No. 3.2;  
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 8 KGGSGGSGSMWHPQFEK 24  
DB 101 KGNVSCPSAMWHPQFEK 117

RESULT 4  
US-09-977-137A-8  
Sequence 8, Application US/09977137A  
Patent No. 6750042  
GENERAL INFORMATION:  
APPLICANT: Summers, Anne O.  
APPLICANT: Caguiat, Jonathan  
TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and  
FILE REFERENCE: 79-00  
CURRENT APPLICATION NUMBER: US/09/977,137A  
PRIOR FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: US 60/240,465  
PRIOR FILING DATE: 2000-10-12  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 8  
LENGTH: 117  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: chelon  
US-09-977-137A-8

Query Match 40.4%; Score 59; DB 4; Length 117;  
Best Local Similarity 64.7%; Pred. No. 3.2;  
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 8 KGGSGGSGSMWHPQFEK 24  
DB 101 KGNVSCPSAMWHPQFEK 117

RESULT 5  
US-09-977-137A-9  
Sequence 9, Application US/09977137A  
Patent No. 6750042  
GENERAL INFORMATION:  
APPLICANT: Summers, Anne O.  
APPLICANT: Caguiat, Jonathan  
TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and  
FILE REFERENCE: 79-00  
CURRENT APPLICATION NUMBER: US/09/977,137A  
PRIOR FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: US 60/240,465  
PRIOR FILING DATE: 2000-10-12  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 9  
LENGTH: 117  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: chelon  
US-09-977-137A-9

Query Match 40.4%; Score 59; DB 4; Length 117;  
Best Local Similarity 64.7%; Pred. No. 3.2;  
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 8 KGGSGGSGSMWHPQFEK 24  
DB 101 KGNVSCPSAMWHPQFEK 117

RESULT 6  
US-09-977-137A-10  
Sequence 10, Application US/09977137A  
Patent No. 6750042  
GENERAL INFORMATION:  
APPLICANT: Summers, Anne O.  
APPLICANT: Caguiat, Jonathan  
TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and  
FILE REFERENCE: 79-00  
CURRENT APPLICATION NUMBER: US/09/977,137A  
PRIOR FILING DATE: 2001-10-12  
PRIOR APPLICATION NUMBER: US 60/240,465  
PRIOR FILING DATE: 2000-10-12  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn Ver. 2.0  
SEQ ID NO 10  
LENGTH: 117  
TYPE: PRT  
ORGANISM: Artificial Sequence  
FEATURE:  
OTHER INFORMATION: Description of Artificial Sequence: chelon  
US-09-977-137A-10

Query Match 40.4%; Score 59; DB 4; Length 117;  
Best Local Similarity 64.7%; Pred. No. 3.2;  
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 8 KGGSGGSGSMWHPQFEK 24  
DB 101 KGNVSCPSAMWHPQFEK 117

```
RESULT 7
US-09-977-137A-12
; Sequence 12, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliati, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-12

Query Match          40.4%; Score 59; DB 4; Length 117;
Best Local Similarity 64.7%; Pred. No. 3.2;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      8 KGGGSGGSMWHPQEK 24
Db      101 KGNVSCPSAWSHQPEK 117

RESULT 8
US-09-977-137A-6
; Sequence 6, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliati, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-6

Query Match          40.4%; Score 59; DB 4; Length 118;
Best Local Similarity 64.7%; Pred. No. 3.2;
Matches 11; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

QY      8 KGGGSGGSMWHPQEK 24
Db      102 KGNVSCPSAWSHQPEK 118

RESULT 9
US-09-407-687-41
; Sequence 41, Application US/09407687
; Patent No. 6548634
; GENERAL INFORMATION:
```

```
; APPLICANT: Ballinger, Marcus
; APPLICANT: Kavanaugh, Michael
; TITLE OF INVENTION: Synthetic Peptides Having FGF Receptor
; TITLE OF INVENTION: Affinity
; FILE REFERENCE: 1517.001
; CURRENT APPLICATION NUMBER: US/09/407,687
; CURRENT FILING DATE: 1999-09-28
; EARLIER APPLICATION NUMBER: 60/102,667
; EARLIER FILING DATE: 1998-09-30
; NUMBER OF SEQ ID NOS: 43
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 109
; TYPE: PRT
; ORGANISM: Synthetic Sequence
; FEATURE:
; OTHER INFORMATION: synthetic peptide
US-09-407-687-41

Query Match          39.7%; Score 58; DB 4; Length 109;
Best Local Similarity 68.8%; Pred. No. 3.9;
Matches 11; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY      1 WSHPQEKGGSGSGS 16
        |||:|||||
Db      24 WSHFRNGFGSGSGS 39

RESULT 10
US-09-809-517A-30
; Sequence 30, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136e1 methods for displaying (poly)peptides/proteins on t
; TITLE OF INVENTION: particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 30
; LENGTH: 21
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30

Query Match          38.4%; Score 56; DB 4; Length 21;
Best Local Similarity 81.8%; Pred. No. 1.2;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

QY      14 GGSWHPQEK 24
        |||||||
Db      11 GAPWHPQEK 21

RESULT 11
US-09-809-517A-33
; Sequence 33, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136e1 methods for displaying (poly)peptides/proteins on t
; TITLE OF INVENTION: particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
```

```

; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 22
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33
```

```

Query Match          38.4%; Score 56; DB 4; Length 22;
Best Local Similarity 81.8%; Pred. No. 1.3;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 14 GGSWSHPQPEK 24
DB 12 GAPWSHPQPEK 22
```

```

RESULT 12
US-09-809-517A-31
; Sequence 31, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136e1 methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
; LENGTH: 24
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-31
```

```

Query Match          38.4%; Score 56; DB 4; Length 24;
Best Local Similarity 81.8%; Pred. No. 1.4;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 14 GGSWSHPQPEK 24
DB 14 GAPWSHPQPEK 24
```

```

RESULT 13
US-09-809-517A-34
; Sequence 34, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136e1 methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
```

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34
```

```

Query Match          38.4%; Score 56; DB 4; Length 25;
Best Local Similarity 81.8%; Pred. No. 1.5;
Matches 9; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY 14 GGSWSHPQPEK 24
DB 15 GAPWSHPQPEK 25
```

```

RESULT 14
US-09-431-887-33
; Sequence 33, Application US/09431887
; Patent No. 6534036
; GENERAL INFORMATION:
; APPLICANT: D-Gen Limited
; TITLE OF INVENTION: BIOLOGICAL MATERIALS AND METHODS USEFUL IN THE
; FILE REFERENCE: ICOT/P21952
; CURRENT APPLICATION NUMBER: US/09/431,887
; PRIOR FILING DATE: 1999-11-02
; PRIOR APPLICATION NUMBER: GB 9824091.4
; PRIOR FILING DATE: 1999-11-04
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 33
; LENGTH: 259
; TYPE: PRT
; ORGANISM: Trichosurus vulpecula
US-09-431-887-33
```

```

Query Match          38.4%; Score 56; DB 4; Length 259;
Best Local Similarity 40.7%; Pred. No. 17;
Matches 11; Conservative 2; Mismatches 4; Indels 10; Gaps 2;
```

```
QY 1 MSHPQPEKGGSS-----GGGSWSHP 20
DB 58 WHPQ-----GGSTWGPQHPGGSNWGP 81
```

```

RESULT 15
US-08-770-761A-2
; Sequence 2, Application US/08770761A
; Patent No. 5814503
; GENERAL INFORMATION:
; APPLICANT: Kovacevic, Steven
; APPLICANT: Rao, Ramachandra N.
; TITLE OF INVENTION: FUSION PROTEINS COMPRISING CELL CYCLE
; TITLE OF INVENTION: REGULATORY PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center/Patent Division
; CITY: Indianapolis
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/770,761A
```

FILED DATE: 19-DEC-1996  
CLASSIFICATION: 530  
ATTORNEY/AGENT INFORMATION:  
NAME: Gaylo, Paul J.  
REGISTRATION NUMBER: 36, 808  
REFERENCE/DOCKET NUMBER: X-10136  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 317-376-0756  
TELEFAX: 317-277-1917  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 660 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
US-08-770-761A-2

Query Match 38.4%; Score 56; DB 2; Length 660;  
Best Local Similarity 50.0%; Pred. No. 44;  
Matches 8; Conservative 3; Mismatches 5; Indels 0; Gaps 0;  
Oy 7 EKGGSGGGSGWSHPQF 22  
Db 643 DEGNPEGGSAMRHPQF 658

Search completed: March 2, 2005, 12:25:36  
Job time : 27.3171 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 18.3415 Seconds  
(without alignments)  
125.901 Million cell updates/sec

Title: SEQ12  
Perfect score: 110  
Sequence: 1 wshpgefxxxxxxwshpgefek 24

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	59	53.6	298	2	D89982	hypothetical prote
2	57	51.8	650	2	T23175	hypothetical prote
3	56.5	51.4	629	2	AF0501	glucose inhibited
4	55	50.0	385	2	D72211	acetylornithine am
5	55	50.0	699	2	A54660	histidine rich cal
6	54	49.1	83	1	C42645	ribosomal protein
7	54	49.1	83	2	F81664	ribosomal protein
8	54	49.1	435	2	A44308	Antho-Rhamide prec
9	53	48.2	147	2	E89838	hypothetical prote
10	53	48.2	268	2	G81305	probable dnaJ-like
11	53	48.2	338	2	H96765	hypothetical prote
12	52	47.3	218	2	AH1117	transaldolase homo
13	52	47.3	218	2	AB1478	transaldolase homo
14	52	47.3	397	2	C84904	hypothetical prote
15	52	47.3	471	2	JQ1021	alpha-galactosidas
16	52	47.3	803	2	T46179	hypothetical prote
17	52	47.3	2128	2	I52577	beta-spectrin - mo
18	51	46.4	86	2	C86570	S17 ribosomal prot
19	51	46.4	86	2	H72054	ribosomal protein
20	51	46.4	124	2	A30043	triphoblast-specif
21	51	46.4	430	2	AH3420	phage DNA packagin
22	51	46.4	458	2	T16041	hypothetical prote
23	51	46.4	466	2	AH1126	endo-1,4-beta-xyla
24	51	46.4	481	2	T32991	hypothetical prote
25	51	46.4	537	1	YRHUB6	tyrosinase-related
26	51	46.4	537	1	YRMSB6	tyrosinase-related
27	51	46.4	547	2	G71307	probable oligopept
28	51	46.4	597	1	HOECT1	hydrogenase (EC 1.
29	51	46.4	597	2	AE0721	hydrogenase-1 larg

30	51	46.4	597	2 B85632	hydrogenase-1 larg
31	51	46.4	597	2 A90770	hydrogenase-1 larg
32	51	46.4	1418	2 T15232	hypothetical prote
33	51	46.4	2469	2 H36812	hypothetical prote
34	50	45.5	103	2 B90973	hypothetical prote
35	50	45.5	103	2 A85821	unknown protein en
36	50	45.5	103	2 C85713	unknown protein en
37	50	45.5	103	2 A90904	hypothetical prote
38	50	45.5	243	2 T08785	hypothetical prote
39	50	45.5	277	2 A97539	MGC1873 protein ho
40	50	45.5	277	2 AD2758	esterase D (import
41	50	45.5	284	2 AH3479	S-formylglutathion
42	50	45.5	310	2 C84701	hypothetical prote
43	50	45.5	334	2 A39172	Antho-Rhamide neur
44	50	45.5	467	2 C87349	efflux system prot
45	50	45.5	534	2 T30268	oligo-1,6-glucosid

## ALIGNMENTS

RESULT 1  
D89982  
hypothetical protein SA1747 [imported] - Staphylococcus aureus (strain N315)  
C:Species: Staphylococcus aureus  
C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004  
C:Accession: D89982  
R:Kuroda, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Oguc  
ma, A.; Mizutani-Uji, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.;  
C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiratsugu, K.  
A:Title: Whole genome sequencing of methicillin-resistant Staphylococcus aureus.  
A:Reference number: A89758; MUID:21311952; PMID:11418146  
A:Accession: D89982  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-298 <KUR>  
A:Cross-references: UNIPROT:Q99SV5; GB:BA000018; PID:G13701726; PIDN:BA843019.1; GSPDB:G  
A:Experimental source: strain N315  
C:Genetics:  
A:Gene: SA1747

Query Match 53.6%; Score 59; DB 2; Length 298;  
Best Local Similarity 34.8%; Pred. No. 4.1;  
Matches 8; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXWSHPQFE 23  
| : |||:::|||||  
Db 90 WTKQLEKMIAPFYRKMDHQVE 112

RESULT 2  
T23175  
hypothetical protein K01C8.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T23175  
R:Sim, M.  
submitted to the EMBL Data Library, April 1995  
A:Reference number: Z19702  
A:Accession: T23175  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-650 <WIL>  
A:Cross-references: UNIPROT:Q21087; EMBL:Z49068; PIDN:CAA88862.1; GSPDB:GN00020; CESP:X0  
A:Experimental source: clone K01C8  
C:Genetics:  
A:Gene: CESP:K01C8.3  
A:Map position: 2  
A:Introns: 31/1; 74/1; 181/3; 392/2; 416/1; 489/3; 551/1  
F:78-546/Domain: animal histidine decarboxylase homology <HDC>  
Query Match 51.8%; Score 57; DB 2; Length 650;





Nucleic Acids Res. 28, 1397-1406, 2000

A>Title: Genome sequences of *Chlamydia trachomatis* Mopn and *Chlamydia pneumoniae* AR39.

A:Reference number: AB1500; MUID:20150235; PMID:10684935

A:Accession: F81664

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-83 <TET>

A:Cross-references: UNIPROT:Q9PUM3; GB:AE002347; GB:AE002160; NID:g7190815; PIDN:AAF3960

A:Experimental source: strain N195 (Mopn)

C:Genetics: TC0806

C:Superfamily: Escherichia coli ribosomal protein S17

Query Match 49.1%; Score 54; DB 2; Length 83;

Best Local Similarity 26.1%; Pred. No. 4.7;

Matches 6; Conservative 14; Mismatches 3; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXSHPOFE 23

Db 33 YSHPOYAKVVRDSKYVAINELD 55

RESULT 8

Antho-Rfam: precursor - sea anemone (*Anthopleura elegantissima*)

C:Species: *Anthopleura elegantissima*

C:Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004

C:Accession: A44308

R:Schmutzler, C.; Darmer, D.; Diekhoff, D.; Grimmelikhuijzen, C.J.

J. Biol. Chem. 267, 22534-22541, 1992

A>Title: Identification of a novel type of processing sites in the precursor for the sea

A:Reference number: A44308; MUID:93054550; PMID:1429603

A:Accession: A44308

A:Status: preliminary

A:Molecule type: mRNA; protein

A:Residues: 1-435 <SCH>

A:Cross-references: UNIPROT:P10419; GB:M98269; NID:g155702; PIDN:AA27738.1; PID:g155703

A>Note: sequence extracted from NCBI backbone (NCBIN:117102, NCBI:P117104)

C:Keywords: neuropeptide

Query Match 49.1%; Score 54; DB 2; Length 435;

Best Local Similarity 39.3%; Pred. No. 32;

Matches 11; Conservative 9; Mismatches 4; Indels 4; Gaps 1;

Qy 1 WSHPOFEKXXXXXXXXXSHPOFE 24

Db 66 FSDPOFWKGRFSDPOFWKGRFSDPOFWK 93

RESULT 9

hypothetical protein [imported] - *Staphylococcus aureus* (strain N315)

C:Species: *Staphylococcus aureus*

C:Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004

C:Accession: E89838

R:Kurita, M.; Ohta, T.; Uchiyama, I.; Baba, T.; Yuzawa, H.; Kobayashi, I.; Cui, L.; Ogino, M.; Mizutani-U, Y.; Kobayashi, N.; Sawano, T.; Inoue, R.; Kaito, C.; Sekimizu, K.; C.; Shiba, T.; Hattori, M.; Ogasawara, N.; Hayashi, H.; Hiramatsu, K.

Lancet 357, 1225-1240, 2001

A>Title: Whole genome sequencing of methicillin-resistant *Staphylococcus aureus*.

A:Reference number: A89758; MUID:21311952; PMID:11410146

A:Accession: E89838

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-147 <KUR>

A:Cross-references: UNIPROT:Q99VU5; GB:BA000018; PID:g13700567; PIDN:BA81864.1; GSPDB:C

A:Experimental source: strain N315

C:Genetics: SA0631

Query Match 48.2%; Score 53; DB 2; Length 147;

Best Local Similarity 33.3%; Pred. No. 13;

Matches 8; Conservative 9; Mismatches 7; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXSHPOFE 24

Db 69 WGFHSEKSMVNIELLYVEPOFRK 92

RESULT 10

probable dnaJ-like protein Cj1034c [imported] - *Campylobacter jejuni* (strain NCTC 11168)

C:Species: *Campylobacter jejuni*

C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004

C:Accession: G81305

R:Parkhill, J.; Wren, B.W.; Mungall, K.; Kelsey, J.M.; Churcher, C.; Basham, D.; Chilling, C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Barral

Nature 403, 665-668, 2000

A>Title: The genome sequence of the food-borne pathogen *Campylobacter jejuni* reveals hyf

A:Reference number: AB1250; MUID:20150912; PMID:10688204

A:Accession: G81305

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-268 <PAR>

A:Cross-references: UNIPROT:Q9PNQ7; GB:ALJ39077; GB:ALJ11168; NID:g6968444; PIDN:CB7325

A:Experimental source: serotype O2, strain NCTC 11168

C:Genetics: Cj1034c

Query Match 48.2%; Score 53; DB 2; Length 268;

Best Local Similarity 31.8%; Pred. No. 25;

Matches 7; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

Qy 3 HPQFEKXXXXXXXXXSHPOFE 24

Db 231 HPDFHQGSALEKAYAREQFEK 252

RESULT 11

hypothetical protein F25P22.26 [imported] - *Arabidopsis thaliana*

C:Species: *Arabidopsis thaliana* (mouse-ear cress)

C:Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004

C:Accession: H96765

R:Theologis, A.; Becker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, J.; Chen, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.R.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A:Authors: Hunter, J.D.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maiti, R.; Marziani, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, K.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.

A>Title: Sequence and analysis of chromosome 1 of the plant *Arabidopsis*.

A:Reference number: AB6141; MUID:21016719; PMID:11130712

A:Accession: H96765

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-338 <STO>

A:Cross-references: UNIPROT:Q9C9S8; GB:AE005173; NID:g6692747; PIDN:AAF24853.1; GSPDB:G

C:Genetics: F25P22.26

A:Map position: 1

Query Match 48.2%; Score 53; DB 2; Length 338;

Best Local Similarity 38.1%; Pred. No. 33;

Matches 8; Conservative 10; Mismatches 3; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXSHPO 21

Db 101 FSHPOGHLPAYTWSMHPQ 121

RESULT 12

transaldolase homolog lmo0343 [imported] - *Listeria monocytogenes* (strain EGD-e)

C:Species: *Listeria monocytogenes*  
 C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
 C:Accession: AH1117  
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, R.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgey, O.; Entlian, K.D.; Fahn, H.; Jones, L.M.; Karet, U.  
 Science 294, 849-852, 2001  
 A:Authors: Krefte, U.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma, A.; Schluter, T.; Simoes, N.; Tillet, A.; Vazquez-Boland, J.A.; Vose, H.; Wehlund, O.; C.; Schluter, T.; Simoes, N.; Tillet, A.; Vazquez-Boland, J.A.; Vose, H.; Wehlund, O.  
 A:Title: Comparative genomics of *Listeria* species.  
 A:Reference number: AB1077; MUID:21537279; PMID:11679669  
 A:Accession: AH1117  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-218 <GLA>  
 A:Cross-references: UNIPROT:Q92EU7; GB:NC\_003210; PIDN:CA98422.1; PID:G16409721; GSPDB: A:Experimental source: strain EGD-e  
 C:Genetics:  
 A:Gene: lmo0343  
 C:Superfamily: Bacillus subtilis 23K phosphoprotein orfU

Query Match 47.3%; Score 52; DB 2; Length 218;  
 Best Local Similarity 35.3%; Pred. No. 27;  
 Matches 6; Conservative 9; Mismatches 2; Indels 0; Gaps 0;

OY 1 WSHPOFEKXXXXXXXXXW 17  
 ||| :|:|:|:|:|:|  
 DB 193 WSHPLTKGIEGFLKDW 209

RESULT 13  
 AB1478  
 transaldolase homolog lmo361 [imported] - *Listeria innocua* (strain Clp11262)  
 C:Species: *Listeria innocua*  
 C>Date: 27-Nov-2001 #sequence\_revision 27-Nov-2001 #text\_change 09-Jul-2004  
 C:Accession: AB1478  
 R:Glaser, P.; Frangeul, L.; Buchrieser, C.; Amend, A.; Baquero, F.; Berche, P.; Bloeker, R.; Dominguez-Bernal, G.; Duchaud, E.; Durand, L.; Dussurgey, O.; Entlian, K.D.; Fahn, H.; Jones, L.M.; Karet, U.  
 Science 294, 849-852, 2001  
 A:Authors: Krefte, U.; Kuhn, M.; Kunst, F.; Kurapat, G.; Madueno, E.; Maitournam, A.; Ma, A.; Schluter, T.; Simoes, N.; Tillet, A.; Vazquez-Boland, J.A.; Vose, H.; Wehlund, O.; C.; Schluter, T.; Simoes, N.; Tillet, A.; Vazquez-Boland, J.A.; Vose, H.; Wehlund, O.  
 A:Title: Comparative genomics of *Listeria* species.  
 A:Reference number: AB1077; MUID:21537279; PMID:11679669  
 A:Accession: AB1478  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-218 <GLA>  
 A:Cross-references: UNIPROT:Q92EU7; GB:AL592022; PIDN:CA95594.1; PID:G16412790; GSPDB: A:Experimental source: strain Clp11262  
 C:Genetics:  
 A:Gene: lmo361  
 C:Superfamily: Bacillus subtilis 23K phosphoprotein orfU

Query Match 47.3%; Score 52; DB 2; Length 218;  
 Best Local Similarity 35.3%; Pred. No. 27;  
 Matches 6; Conservative 9; Mismatches 2; Indels 0; Gaps 0;

OY 1 WSHPOFEKXXXXXXXXXW 17  
 ||| :|:|:|:|:|:|  
 DB 193 WSHPLTKGIEGFLKDW 209

RESULT 14  
 C84904  
 hypothetical protein At2g46550 [imported] - *Arabidopsis thaliana*  
 C:Species: *Arabidopsis thaliana* (mouse-ear cress)  
 C>Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004  
 C:Accession: C84904  
 R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.; eues, D.; Nierman, W.C.; White, O.; Eilen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, J.  
 Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.  
 A:Reference number: A84420; MUID:20083487; PMID:10617157  
 A:Accession: C84904  
 A:Status: preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-397 <STO>  
 A:Cross-references: UNIPROT:Q92EP4; GB:AE002093; NID:G4415936; PIDN:AA20166.1; GSPDB:GN  
 C:Genetics:  
 A:Gene: At2g46550  
 A:Map position: 2

Query Match 47.3%; Score 52; DB 2; Length 397;  
 Best Local Similarity 34.8%; Pred. No. 54;  
 Matches 8; Conservative 10; Mismatches 3; Indels 2; Gaps 1;

OY 3 HPQFEKXXXXXXXXXWSP -QFE 23  
 ||| :|:|:|:|:|:|  
 DB 74 NPNFPLDLPNTRLMSPHPOFO 96

RESULT 15  
 JQ1021  
 alpha-galactosidase (EC 3.2.1.22) - Yeast (*Saccharomyces cerevisiae*) (strain carlsberg  
 C:Species: *Saccharomyces cerevisiae*  
 A:Variety: *Saccharomyces carlsbergensis* NCYC396  
 C>Date: 17-Jul-1992 #sequence\_revision 17-Jul-1992 #text\_change 09-Jul-2004  
 C:Accession: JQ1021  
 R:Turakainen, H.; Korhola, M.; Aho, S.  
 Gene 101, 97-104, 1991  
 A:Title: Cloning, sequence and chromosomal location of a MEL gene from *Saccharomyces car*  
 A:Reference number: JQ1021; MUID:91285441; PMID:1711992  
 A:Accession: JQ1021  
 A:Molecule type: DNA  
 A:Residues: 1-471 <TUR>  
 A:Cross-references: UNIPROT:Q03647; GB:M58484; NID:G171923; PIDN:AAA4769.1; PID:G171924  
 A:Experimental source: Strain NCYC396  
 A:Note: the source is designated as *Saccharomyces carlsbergensis*  
 C:Comment: Yeast strains producing this enzyme are able to use melibiose as a carbon sou  
 C:Genetics:  
 A:Gene: MEL  
 A:Map position: 10  
 C:Superfamily: alpha-galactosidase  
 C:Keywords: glycosidase; hydrolase

Query Match 47.3%; Score 52; DB 2; Length 471;  
 Best Local Similarity 36.4%; Pred. No. 66;  
 Matches 8; Conservative 9; Mismatches 3; Indels 2; Gaps 1;

OY 1 WSH--POFEKXXXXXXXXXWSP 20  
 ||| :|:|:|:|:|:|  
 DB 328 WRHQPVDTKTGQGIQFWSGP 349

Search completed: March 2, 2005, 12:28:58  
 Job time : 19.3415 secs

Wed Mar 2 14:26:53 2005

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 89.1707 Seconds

(without alignments)  
137.824 Million cell updates/sec

Title: SEQ12  
Perfect score: 110  
Sequence: 1 webpfekxxxxxxxxwebpfek 24

Scoring table: BLOSUM62X  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues  
Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot\_03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60	54.5	743	2 08CH90	08CH90 rattus norv
2	59	53.6	298	2 08NVR8	08NVR8 staphylococ
3	59	53.6	298	2 09SV5	09SV5 staphylococ
4	59	53.6	298	2 07A4N3	07A4N3 staphylococ
5	59	53.6	298	2 06G800	06G800 staphylococ
6	59	53.6	298	2 06GFC4	06GFC4 staphylococ
7	59	53.6	508	1 MATK MARSC	08MK9 marathrum s
8	59	53.6	508	2 09BBG0	09BBG0 varroynelli
9	59	53.6	508	2 09BBG3	09BBG3 oserya coul
10	59	53.6	704	2 08CUD2	08CUD2 rattus norv
11	57	51.8	452	2 09H7M8	09H7M8 homo sapien
12	57	51.8	504	2 08BYP3	08BYP3 pseudomonas
13	57	51.8	650	2 021087	021087 caenorhabdi
14	57	51.8	691	2 08ENR9	08ENR9 oceanobacti
15	57	51.8	705	2 095ZS2	095ZS2 caenorhabdi
16	57	51.8	1714	2 09FMM3	09FMM3 arabidopsis
17	56.5	51.4	107	2 06ZJ15	06ZJ15 oryza sativ
18	56.5	51.4	629	1 GIDA YERPE	08Z98 yerinia pe
19	56.5	51.4	629	2 0663P9	0663P9 yerinia pe
20	55.5	50.5	389	2 09AD61	09AD61 streptomyce
21	55	50.0	256	2 06GZX4	06GZX4 frog vitus
22	55	50.0	355	2 09T1Z2	09T1Z2 bacterioph
23	55	50.0	385	1 ARGD THEME	09X25 thermocoga
24	55	50.0	436	2 08BRY1	08BRY1 lactobacill
25	55	50.0	699	1 SRCH HUMAN	08B18 pseudomonas
26	54.5	49.5	1131	2 08B1I8	08B1I8 pseudomonas
27	54	49.1	83	1 RS17_CHLNU	09PJM3 chlamydia m
28	54	49.1	83	1 RS17_CHLTR	08B18 pseudomonas
29	54	49.1	208	2 062TR3	062TR3 bacillus th
30	54	49.1	256	2 065IB2	065IB2 bacillus th
31	54	49.1	265	2 09PWS6	09PWS6 bothrops ja

32	54	49.1	429	1 FMR2 ANTEL	016994 anthopleura
33	54	49.1	435	1 FMR1 ANTEL	P10419 anthopleura
34	54	49.1	512	2 091402	091402 polyangium
35	54	49.1	535	1 TYR1 CHICK	057405 gallus gall
36	54	49.1	536	2 091BH8	091BH8 coturnix co
37	54	49.1	587	2 08M232	08M232 dirosophila
38	54	49.1	587	2 09V9B8	09V9B8 dirosophila
39	54	49.1	802	2 07UMP8	07UMP8 rhodospirilla
40	54	49.1	1676	2 08G36	08G36 utillago ma
41	53.5	48.6	407	2 06NGV4	06NGV4 corynebacte
42	53.5	48.6	848	2 08ZRR6	08ZRR6 salmoneila
43	53	48.2	147	2 08NXQ8	08NXQ8 staphylococ
44	53	48.2	147	2 093VU5	093VU5 staphylococ
45	53	48.2	147	2 07A6X9	07A6X9 staphylococ

## ALIGNMENTS

```

RESULT 1
ID 08CH90 PRELIMINARY; PRT; 743 AA.
AC 08CH90;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DE Soluble guanylyl cyclase alpha 2 subunit E219G mutant.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Nakamura I., Yao Y., Suzuki N.;
RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB097860; BAC44887.1; -.
DR HSSP; P30803; IAZS.
DR GO; GO:0004383; P-guanylate cyclase activity; IEA.
DR GO; GO:0016829; P-lyase activity; IEA.
DR GO; GO:0007242; P-intracellular signaling cascade; IEA.
DR InterPro; IPR001054; G_cyclase.
DR InterPro; IPR009080; tRNAsyn Ia bind.
DR Pfam; PF00211; Guanylate_cyc; 1.
DR SMART; SM00044; CYCC; 1.
DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR PROSITE; PS00125; GUANYLATE_CYCLASES_2; 1.
KW Lyase.
SQ SEQUENCE 743 AA; 83251 MW; ACF5C53E0982813A CRC64;

Query Match 54.5%; Score 60; DB 2; Length 743;
Best Local Similarity 47.1%; Pred. No. 40;
Matches 8; Conservative 9; Mismatches 0; Indels 0; Gaps 0;

QY 8 KXXXXXXXXXSHPOFEK 24
Db 727 ETSLVPRGSMWHPFEK 743

RESULT 2
ID 08NVR8 PRELIMINARY; PRT; 298 AA.
AC 08NVR8;
DT 01-OCT-2002 (TREMBlrel. 22, Created)
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)
DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE MW1874 protein.
GN Ordered.cusNames=MW1874;
OS Staphylococcus aureus (strain MW2).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_Taxid=196620;
RN [1]
RP SEQUENCE FROM N.A.

```

```
RC STRAIN=MM2;
RX MEDLINE=22040717; PubMed=12044378; DOI=10.1016/S0140-6736(02)08713-5;
RA Baba T., Takeuchi F., Kuroda M., Yuzawa H., Aoki K.-I., Oguchi A.,
RA Nagai Y., Traama N., Amano K., Naimi T., Kuroda H., Cui L.,
RA Yamamoto K., Hiramatsu K.;
RT "Genome and virulence determinants of high virulence community-
acquired MRSA."
RL Lancet 359:1819-1827(2002).
CC -1- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL; AP003482; BAB95739.1;
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO; GO:0003684; F:damaged DNA binding; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006298; P:mismatch repair; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD001263; MutS_C; 2.
DR ProDom; PD000006; ABC_tran; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR ATP-binding; Complete proteome.
SQ SEQUENCE 298 AA; 34532 MW; 0DC7413478EA7948 CRC64;

Query Match 53.6%; Score 59; DB 2; Length 298;
Best Local Similarity 34.8%; Pred. No. 24;
Matches 8; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXSHPOFE 23
Db 90 WTKOLEKMIAPFYRKMDHQVFE 112

RESULT 3
099SV5 PRELIMINARY; PRT; 298 AA.
ID 099SV5;
AC 099SV5;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Similar to ABC transporter (ATP-binding protein).
GN OrderedLocustNames=SA1747;
OS Staphylococcus aureus (strain Mu50 / ATCC 700699).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158878;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21311952; PubMed=11418146; DOI=10.1016/S0140-6736(00)04403-2;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hasegawa A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hiraoka H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hatori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus."
RL Lancet 357:1225-1240(2001).
CC -1- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL; AP003363; BAB58095.1; -.
DR PIR; D89982; D89982.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO; GO:0003684; F:damaged DNA binding; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006298; P:mismatch repair; IEA.
DR GO; GO:0006810; P:transport; IEA.
```

```
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR InterPro; IPR000432; MutS_C.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR ProDom; PD001263; MutS_C; 2.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR ATP-binding; Complete proteome.
SQ SEQUENCE 298 AA; 34439 MW; 6FAAE51CB048B334 CRC64;

Query Match 53.6%; Score 59; DB 2; Length 298;
Best Local Similarity 34.8%; Pred. No. 24;
Matches 8; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXSHPOFE 23
Db 90 WTKOLEKMIAPFYRKMDHQVFE 112

RESULT 4
07A4N3 PRELIMINARY; PRT; 298 AA.
ID 07A4N3;
AC 07A4N3;
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE SA1747 protein.
GN OrderedLocustNames=SA1747;
OS Staphylococcus aureus (strain N315).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
OX NCBI_TaxID=158879;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21311952; PubMed=11418146; DOI=10.1016/S0140-6736(00)04403-2;
RA Kuroda M., Ohta T., Uchiyama I., Baba T., Yuzawa H., Kobayashi I.,
RA Cui L., Oguchi A., Aoki K.-I., Nagai Y., Iian J.-Q., Ito T.,
RA Kanamori M., Matsumaru H., Maruyama A., Murakami H., Hasegawa A.,
RA Mizutani-Ui Y., Takahashi N.K., Sawano T., Inoue R.-I., Kaito C.,
RA Sekimizu K., Hiraoka H., Kuhara S., Goto S., Yabuzaki J.,
RA Kanehisa M., Yamashita A., Oshima K., Furuya K., Yoshino C., Shiba T.,
RA Hatori M., Ogasawara N., Hayashi H., Hiramatsu K.;
RT "Whole genome sequencing of methicillin-resistant Staphylococcus
aureus."
RL Lancet 357:1225-1240(2001).
CC -1- SIMILARITY: Belongs to the ABC transporter family.
DR EMBL; AP003135; BAB43019.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0005524; F:ATP binding; IEA.
DR GO; GO:0042626; F:ATPase activity, coupled to transmembrane m. . .; IEA.
DR GO; GO:0003684; F:damaged DNA binding; IEA.
DR GO; GO:0000166; F:nucleotide binding; IEA.
DR GO; GO:0006298; P:mismatch repair; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003593; AAA_ATPase.
DR InterPro; IPR003439; ABC_transporter.
DR Pfam; PF00005; ABC_tran; 1.
DR ProDom; PD000006; ABC_transporter; 1.
DR ProDom; PD001263; MutS_C; 2.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR ATP-binding; Complete proteome.
SQ SEQUENCE 298 AA; 34439 MW; 6FAAE51CB048B334 CRC64;

Query Match 53.6%; Score 59; DB 2; Length 298;
Best Local Similarity 34.8%; Pred. No. 24;
Matches 8; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXSHPOFE 23
Db 90 WTKOLEKMIAPFYRKMDHQVFE 112
```

```

RESULT 5
OG6800 PRELIMINARY; PRT; 298 AA.
AC OG6800;
DT 05-JUL-2004 (TEMBLrel. 27, Created)
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)
DE Putative ABC transporter ATP-binding protein.
GN OrderedLocuNames=SA51857;
OC Staphylococcus aureus (strain MSSA476).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
NCBI_TaxID=282459;
RN [1]
RX PUBLISHED FROM N.A.
  PubMed=15233324; DOI=10.1073/pnas.0402521101;
  Holden M.T.G., Fell E.J., Lindsey J.A., Peacock S.J., Day N.P.J.,
  Enright M.C., Foster T.J., Moore C.E., Hurst L., Atkin R., Barron A.,
  James K.D., Lennard N., Line A., Mayes R., Moule S., Mungall K.,
  Orchard D., Quail M.A., Rabinowitch E., Rutherford K.M., Sanders M.,
  Sharp S., Simmonds M., Stevens K., Whitehead S., Barrett B.G.,
  Spratt B.G., Parkhill J.;
  "Complete genomes of two clinical Staphylococcus aureus strains:
  evidence for the rapid evolution of virulence and drug resistance.";
  Proc. Natl. Acad. Sci. U.S.A. 101:9786-9791(2004).
  -1- SIMILARITY: Belongs to the ABC transporter family.
  EMBL; BX571857; CAG4362.1; -.
DR GO:GO:0016020; C:membrane; IEA.
DR GO:GO:0005524; F:ATP binding; IEA.
DR GO:GO:0042626; F:ATPase activity, coupled to transmembrane m. .; IEA.
DR GO:GO:0003684; F:damaged DNA binding; IEA.
DR GO:GO:0000166; F:nucleotide binding; IEA.
DR GO:GO:0006298; P:mismatch repair; IEA.
DR GO:GO:0006810; P:transport; IEA.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran; I.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR ATP-binding; Complete proteome.
SQ SEQUENCE 298 AA; 34532 MW; 0DC7413478EA7948 CRC64;

Query Match 53.6%; Score 59; DB 2; Length 298;
Best Local Similarity 34.8%; Pred. No. 24;
Matches 8; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

OY 1 WSHPOFEKXXXXXXXXXSHPOFE 23
   | | | | | | | | | | | | | |
DB 90 WTKQLEKMIAPFYRKMDHVF 112

RESULT 6
OG6FC4 PRELIMINARY; PRT; 298 AA.
AC OG6FC4;
DT 05-JUL-2004 (TEMBLrel. 27, Created)
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)
DE Putative ABC transporter ATP-binding protein.
GN OrderedLocuNames=SA82025;
OC Staphylococcus aureus (strain MRSA252).
OC Bacteria; Firmicutes; Bacillales; Staphylococcus.
NCBI_TaxID=282458;
RN [1]
RX PUBLISHED FROM N.A.
  PubMed=1521324; DOI=10.1073/pnas.0402521101;
  Holden M.T.G., Fell E.J., Lindsey J.A., Peacock S.J., Day N.P.J.,

```

```

  Enright M.C., Foster T.J., Moore C.E., Hurst L., Atkin R., Barron A.,
  Barron N., Bentley S.D., Chillingworth C., Chillingworth T.,
  Churcher C., Clark L., Corton C., Cronin A., Doggett J., Dowd L.,
  Felwell T., Hance Z., Harris B., Hauser H., Holroyd S., Jagels K.,
  James K.D., Lennard N., Line A., Mayes R., Moule S., Mungall K.,
  Orchard D., Quail M.A., Rabinowitch E., Rutherford K.M., Sanders M.,
  Sharp S., Simmonds M., Stevens K., Whitehead S., Barrett B.G.,
  Spratt B.G., Parkhill J.;
  "Complete genomes of two clinical Staphylococcus aureus strains:
  evidence for the rapid evolution of virulence and drug resistance.";
  Proc. Natl. Acad. Sci. U.S.A. 101:9786-9791(2004).
  -1- SIMILARITY: Belongs to the ABC transporter family.
  EMBL; BX571856; CAG41010.1; -.
DR GO:GO:0016020; C:membrane; IEA.
DR GO:GO:0005524; F:ATP binding; IEA.
DR GO:GO:0042626; F:ATPase activity, coupled to transmembrane m. .; IEA.
DR GO:GO:0003684; F:damaged DNA binding; IEA.
DR GO:GO:0000166; F:nucleotide binding; IEA.
DR GO:GO:0006298; P:mismatch repair; IEA.
DR GO:GO:0006810; P:transport; IEA.
DR InterPro: IPR003593; AAA_ATPase.
DR InterPro: IPR003439; ABC_transporter.
DR Pfam: PF00005; ABC_tran; I.
DR ProDom: PD000006; ABC_transporter; 1.
DR SMART; SM00382; AAA; 1.
DR PROSITE; PS50893; ABC_TRANSPORTER_2; 1.
DR ATP-binding; Complete proteome.
SQ SEQUENCE 298 AA; 34474 MW; 666691085CABCA58 CRC64;

Query Match 53.6%; Score 59; DB 2; Length 298;
Best Local Similarity 34.8%; Pred. No. 24;
Matches 8; Conservative 9; Mismatches 6; Indels 0; Gaps 0;

OY 1 WSHPOFEKXXXXXXXXXSHPOFE 23
   | | | | | | | | | | | | | |
DB 90 WTKQLEKMIAPFYRKMDHVF 112

RESULT 7
MATK_MARSC STANDARD; PRT; 508 AA.
AC O8WK9;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 05-JUL-2004 (Rel. 44, Last annotation update)
DE Maturase K (intron maturase).
GN Name=matK;
OS Marathrum schleddeanum.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid I; Malpighiales; Podostemaceae; Marathrum.
NCBI_TaxID=116737;
RN [1]
RX PUBLISHED FROM N.A.
  Kita Y., Kato M.;
  "Phylogenetic relationships of the aquatic angiosperm family
  Podostemaceae inferred from matK sequence data.";
  Submitted (FEB-2000) to the EMBL/GenBank/DBS databases.
  -1- FUNCTION: Probably assists in splicing chloroplast group II
  introns (by similarity).
  -1- SIMILARITY: Belongs to the intron maturase family 2. MatK
  subfamily.
  =====
  CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
  CC the European Bioinformatics Institute. There are no restrictions on its
  CC use by non-profit institutions as long as its content is in no way
  CC modified and this statement is not removed. Usage by and for commercial
  CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
  CC or send an email to license@isb-sib.ch).

```

```

CC -----
DR EMBL: AB038195; BAB3156.1; -.
DR InterPro: IPR008998; Agglutinin.
DR InterPro: IPR000442; Intron_maturase2.
DR InterPro: IPR002866; MatK_N.
DR Pfam: PF01348; Intron_maturase2; 1.
DR Pfam: PF01824; MatK_N; 1.
DR Chloroplast; mRNA processing.
SQ SEQUENCE 508 AA; 61342 MW; ECCPSBA16B0ABSC3 CRC64;

Query Match
Best Local Similarity 42.1%; Score 59; DB 1; Length 508;
Matches 8; Conservative 9; Mismatches 2; Indels 0; Gaps 0;

QY 6 FEKXXXXXXXXXSHPOFEK 24
   | : : : : : : : : : : : : : :
DB 177 FLRLPFHEYWNWSPHSFEK 195

RESULT 8
Q9BGC0 PRELIMINARY; PRT; 508 AA.
AC Q9BGC0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Maturase K.
GN Name: Maturase K.
OS Vantroyella plumosa.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC euroside I; Malpighiales; Podostemaceae; Vantroyella.
OX NCBI_TaxID=51609;
RN [1]_TaxID=51609;
RP SEQUENCE FROM N.A.
RA Kita Y., Kato M.;
RT "Infrafamilial Phylogeny of the Aquatic Angiosperm Podostemaceae
RT Inferred from the Nucleotide Sequences of the matK Gene.";
RL Plant Biol. 3:156-163(2001).
DR EMBL: AB048378; BAB3398.1; -.
DR GO: GO:0009507; C:chloroplast; IEA.
DR GO: GO:0008380; P:RNA splicing; IEA.
DR InterPro: IPR008998; Agglutinin.
DR InterPro: IPR000442; Intron_maturase2.
DR InterPro: IPR002866; MatK_N.
DR Pfam: PF01348; Intron_maturase2; 1.
DR Pfam: PF01824; MatK_N; 1.
DR Chloroplast.
KW Chloroplast.
SQ SEQUENCE 508 AA; 61288 MW; 90229C45138AF4B6 CRC64;

Query Match
Best Local Similarity 42.1%; Score 59; DB 2; Length 508;
Matches 8; Conservative 9; Mismatches 2; Indels 0; Gaps 0;

QY 6 FEKXXXXXXXXXSHPOFEK 24
   | : : : : : : : : : : : : : :
DB 177 FLRLPFHEYWNWSPHSFEK 195

RESULT 9
Q9BGC3 PRELIMINARY; PRT; 508 AA.
AC Q9BGC3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Maturase K.
GN Name: Maturase K.
OS Oseya coultieriana.
OG Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

```

```

OC euroside I; Malpighiales; Podostemaceae; Oseya.
OX NCBI_TaxID=51602;
RN [1]_TaxID=51602;
RP SEQUENCE FROM N.A.
RA Kita Y., Kato M.;
RT "Infrafamilial Phylogeny of the Aquatic Angiosperm Podostemaceae
RT Inferred from the Nucleotide Sequences of the matK Gene.";
RL Plant Biol. 3:156-163(2001).
DR EMBL: AB048375; BAB3395.1; -.
DR GO: GO:0009507; C:chloroplast; IEA.
DR GO: GO:0008380; P:RNA splicing; IEA.
DR InterPro: IPR008998; Agglutinin.
DR InterPro: IPR000442; Intron_maturase2.
DR InterPro: IPR002866; MatK_N.
DR Pfam: PF01348; Intron_maturase2; 1.
DR Pfam: PF01824; MatK_N; 1.
DR Chloroplast.
KW Chloroplast.
SQ SEQUENCE 508 AA; 61400 MW; 45240CD31B54CDF CRC64;

Query Match
Best Local Similarity 42.1%; Score 59; DB 2; Length 508;
Matches 8; Conservative 9; Mismatches 2; Indels 0; Gaps 0;

QY 6 FEKXXXXXXXXXSHPOFEK 24
   | : : : : : : : : : : : : : :
DB 177 FLRLPFHEYWNWSPHSFEK 195

RESULT 10
Q8CJD2 PRELIMINARY; PRT; 704 AA.
AC Q8CJD2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE Guanylyl cyclase alpha 1 subunit.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]_TaxID=10116;
RP SEQUENCE FROM N.A.
RA Nakamura I., Suzuki N.;
RT Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL: AB096020; BAC24016.1; -.
DR HSPF: P30803; IAZS.
DR GO: GO:0004383; F:guanylate cyclase activity; IEA.
DR GO: GO:0016829; F:lyase activity; IEA.
DR GO: GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro: IPR001054; G_cyclase.
DR Pfam: PF00211; Guanylate_cyc; 1.
DR SMART: SM00044; CYC; 1.
DR PROSITE: PS00452; GUANYLATE_CYCLASES_1; 1.
DR PROSITE: PS50125; GUANYLATE_CYCLASES_2; 1.
KW Lyase.
SQ SEQUENCE 704 AA; 79188 MW; F73FC97B68544E6A CRC64;

Query Match
Best Local Similarity 50.0%; Score 59; DB 2; Length 704;
Matches 8; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY 9 XXXXXXXXXXXXSHPOFEK 24
   | : : : : : : : : : : : : : :
DB 689 VDLVPRGSMWSPHSFEK 704

RESULT 11
Q9H7M8 PRELIMINARY; PRT; 452 AA.
AC Q9H7M8;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

```

```

DE FLJ00042 protein (Fragment).
GN Name=FLJ00042;
OS Homo sapiens (human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
NCBI_TaxID=9606;
[1]
RN
RP
RS
RC TISSUE=Esplein;
RX PubMed:14702039; DOI=10.1038/ng1285;
RA Oca T., Suzuki Y., Nishikawa T., Otsuki T., Sugiyama T., Irie R.,
RA Wakamatsu A., Hayashi K., Sato H., Nagai K., Kimura K., Maki H.,
RA Sekine M., Obayashi M., Nishi T., Shibahara T., Tanaka T., Ishii S.,
RA Yamamoto J., Saito K., Kawai Y., Isono Y., Nakamura Y., Nagahara K.,
RA Murakami K., Yasuda T., Iwayanagi T., Wagatsuma M., Shiratori A.,
RA Sudo H., Hoshino T., Kaku Y., Kodaïra H., Kondo H., Sugawara M.,
RA Takahashi M., Kanda K., Yokoi T., Furuya T., Kikkawa E., Omura Y.,
RA Abe K., Kamihara K., Katsuma N., Sato K., Tanikawa M., Yamazaki M.,
RA Nishimura K., Ishihashi T., Yamashita H., Murakawa M., Fujimori K.,
RA Tanai H., Kimura M., Watanabe M., Hirooka S., Chiba Y., Iehida S.,
RA Ono Y., Takiguchi S., Watanabe S., Yoshida W., Hoshida T., Kusano J.,
RA Kanehori K., Takahashi-Fujii A., Hara H., Tanase T., Nomura Y.,
RA Musashino K., Yuki H., Oshima A., Sasaki N., Aotsuma S.,
RA Yoshikawa Y., Matsunawa H., Ichihara T., Shiohara N., Sano S.,
RA Moriya S., Momiyama H., Satoh N., Takami S., Terashima Y., Suzuki O.,
RA Nakagawa S., Senoh A., Mizoguchi H., Goto Y., Shimizu F., Wakebe H.,
RA Hishigaki H., Watanabe T., Sugiyama A., Takemoto M., Kawakami B.,
RA Yamazaki M., Watanabe K., Kumagai A., Itakura S., Fukuzumi Y.,
RA Fujimori Y., Komiyama M., Tashiro A., Tanigami A., Fujizumi T.,
RA Ono T., Yamada K., Fujii Y., Ozaki K., Hirao M., Omori Y., Okamoto S.,
RA Kawabata A., Hikiji T., Kobatake N., Inagaki H., Ikema Y., Okamoto S.,
RA Okitani R., Kawakami T., Noguchi S., Itoh T., Shigeta K., Senba T.,
RA Matsumura K., Nakajima Y., Mizuno T., Morinaga M., Sasaki M.,
RA Togashi T., Oyama M., Hara H., Watanabe M., Komatsu T.,
RA Mizushima-Sugano J., Sato T., Shirai Y., Takahashi Y., Nakagawa K.,
RA Okumura K., Nagase T., Nomura N., Kikuchi H., Maeno Y., Yamashita R.,
RA Nakai K., Yada T., Nakamura Y., Ohara O., Isegaki T., Sugano S.,
RT "Complete sequencing and characterization of 21,243 full-length human
RT cDNAs."
DR EMBL; AK024450; BAB15740.1; -.
DR GO; GO:0005525; F:GTP binding; IEA.
DR GO; GO:0007364; P:small GTPase mediated signal transduction; IEA.
DR InterPro; IPR001806; Rsg_cnaftmng.
DR PRINTS; PR00449; RASTRNSFRMNG.
FT NON TER
SQ
SEQUENCE 452 AA; 48320 MW; FD10D8A4EA1652B7 CRC64;

Query Match 51.8%; Score 57; DB 2; Length 452;
Best Local Similarity 34.8%; Pred. No. 72;
Matches 8; Conservative 8; Mismatches 7; Indels 0; Gaps 0;

OY 1 WSHPOPEKXXXXXXXXXWSHPOFE 23
Db 77 WSHPALSCDPTSPERWGVQSE 99

```

```

RX MEDLINE=22423060; PubMed=12534463;
RA Nelson K.B., Weibel C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.E., Gill S.R., Pop M., Holmes M.,
RA Brinkac L.M., Beaman M.J., Debey R.T., Daugherty S.C., Kolonay J.F.,
RA Madupu R., Nelson W.C., White O., Peterson J.D., Khouri H.M.,
RA Hance I., Chris Lee P., Holtzapple E.K., Scanlan D., Tran K.,
RA Morazzz A., Utechtack T.R., Rizzo M., Lee K., Kosack D., Moestl D.,
RA Medler H., Lauber J., Stjepandic D., Hohelsel J., Streletz M., Helm S.,
RA Kiewitz C., Eissen J.A., Timmis K.N., Duesterhoeft A., Tuenmler B.,
RA Frazer C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440."
RT Environ. Microbiol. 4:799-808(2002).
DR EMBL; AE016776; AA065340.1; -.
DR TIGR; PE0715; -.
DR GO; GO:0005215; P:transporter activity; IEA.
DR GO; GO:0006810; P:transport; IEA.
DR InterPro; IPR003423; OEP.
DR InterPro; IPR010131; RND_outter_NodT.
DR Pfam; PF02321; OEP; 2. RND_outter_NodT.
DR TIGRPFAMs; TIGR01845; RND_outter_NodT; 1.
KW Complete proteome.
SQ
SEQUENCE 504 AA; 55514 MW; 581092F1804FC14D CRC64;

Query Match 51.8%; Score 57; DB 2; Length 504;
Best Local Similarity 38.1%; Pred. No. 81;
Matches 8; Conservative 9; Mismatches 4; Indels 0; Gaps 0;

OY 4 POFKXXXXXXXXXWSHPOFEK 24
Db 22 POFKQAPWLDGKSTPMIEQ 42

RESULT 13
ID Q21087 PRELIMINARY; PRT; 650 AA.
AC Q21087;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-JAN-1999 (TrEMBLrel. 09, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein K01C8.3a.
GN Name=tdc-1; Synonyms=K01C8.3a; ORFNames=K01C8.3;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Felodierinae; Caenorhabditis.
NCBI_TaxID=6239;
[1]
RN
RP
RS
RC STRAIN=Bristol N2.
RX MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C. elegans: A platform for
RT investigating biology."
RT Science 282:2012-2018 (1998).
RN [2]
RP
RS
RC STRAIN=Bristol N2.
RX Sime M.A.;
RT Submitted (APR-1995) to the EMBL/Genbank/DBJ databases.
CC -1- COFACTOR: Pyridoxal phosphate (By similarity).
CC -1- SIMILARITY: belongs to the group II decarboxylase family.
DR EMBL; Z49068; CA88862.1; -.
DR FIR; T23175; T23175.
DR HSP; P80041; J1631.
DR Wormbase; WBGene0000562; tdc-1.
DR GO; GO:0016831; P:carboxy-lyase activity; IEA.
DR GO; GO:0016823; P:lyase activity; IEA.
DR GO; GO:0006520; P:amino acid metabolism; IEA.
DR InterPro; IPR010977; Pyridoxal_dec.
DR InterPro; IPR002129; Pyridoxal_dec.
DR Pfam; PF00282; Pyridoxal_dec; 1.
DR PRINTS; PR00800; YHDCRBOXLASE.

```

DR PROSITE; PS00392; DDC GAD HDC YDC; 1.  
 RC STRAIN=Bristol N2;  
 KW Hypothetical protein; Lyase; Pyridoxal phosphate.  
 SQ SEQUENCE 650 AA; 73201 MW; 4830ECCAPD1AFAB1 CRC64;

Query Match 51.8%; Score 57; DB 2; Length 650;  
 Best Local Similarity 41.2%; Pred. No. 1.1e+02;  
 Matches 7; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

QY 6 FEKXXXXXXWSHPQF 22  
 DB 136 FEKLIMPGITHMQHPRF 152

## RESULT 14

OEBNA9 PRELIMINARY; PRT; 691 AA.

AC OEBNA9; PRELIMINARY; PRT; 691 AA.  
 DT 01-MAR-2003 (T-EMBLrel. 23, Created)  
 DT 01-MAR-2003 (T-EMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (T-EMBLrel. 25, Last annotation update)  
 DE Hypothetical conserved protein.  
 GN OrderedLocustNames=OB2578;  
 OS Oceanobacillus ihayensis.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Oceanobacillus.  
 CX NCBI\_TaxID=182710;

QY 11 SEQUENCE FROM N.A.

RC STRAIN=HTE831;  
 RX MEDLINE=22220767; PubMed=12235376; DOI=10.1093/nar/gkf526;  
 RA Takami H., Takaki Y., Uchiyama I.;  
 RT "Genome sequence of Oceanobacillus ihayensis isolated from the Iheya  
 RT Ridge and its unexpected adaptive capabilities to extreme  
 RT environments.";  
 RL Nucleic Acids Res. 30:3927-3935(2002).  
 DR EMBL; AP004601; BAC14534.1.  
 DR InterPro; IPR004879; DUF255.  
 DR InterPro; IPR008928; Glyco\_trans\_6np.  
 DR Pfam; PF03190; DUF255; 1.  
 DR Complete proteome; Hypothetical protein.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 691 AA; 79588 MW; 11EP02E245B06B5 CRC64;

Query Match 51.8%; Score 57; DB 2; Length 691;  
 Best Local Similarity 33.3%; Pred. No. 1.2e+02;  
 Matches 8; Conservative 8; Mismatches 8; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXWSHPQF 24  
 DB 247 WDHVGVGRVATDRKMWPPHPEK 270

## RESULT 15

O95ZS2 PRELIMINARY; PRT; 705 AA.

AC O95ZS2;  
 DT 01-DEC-2001 (T-EMBLrel. 19, Created)  
 DT 01-DEC-2001 (T-EMBLrel. 19, Last sequence update)  
 DT 01-MAR-2004 (T-EMBLrel. 26, Last annotation update)  
 DE Hypothetical protein K01C8.3b;  
 GN Name=dc-1; Synonym=K01C8.3b; ORFNames=K01C8.3;  
 OS Caenorhabditis elegans.  
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae;  
 OC Rhabditidae; Peloderae; Caenorhabditis.  
 CX NCBI\_TaxID=6239;

QY 11 SEQUENCE FROM N.A.

RC STRAIN=Bristol N2;  
 RX MEDLINE=99069613; PubMed=9851916;  
 RA none;  
 RT "Genome sequence of the nematode C. elegans: A platform for  
 RT investigating biology.";  
 RL Science 282:2012-2018(1998).  
 RN [2]  
 RP SEQUENCE FROM N.A.

RA Sime M.A.;  
 RL Submitted (APR-1995) to the EMBL/GenBank/DBJ databases.  
 CC -1- CORACOR: Pyridoxal phosphate (By similarity).  
 CC -1- SIMILARITY: Belongs to the group II decarboxylase family.

DR EMBL; Z49068; CAC42319.1; -.  
 DR PIR; T23168; T23168.  
 DR HSSP; P80041; JUS3.  
 DR WormBase; WBGene00006562; tdc-1.  
 DR GO; GO:0016831; F:carboxy-lyase activity; IEA.  
 DR GO; GO:0016829; F:lyase activity; IEA.  
 DR GO; GO:0005520; P:amino acid metabolism; IEA.  
 DR InterPro; IPR010977; Aromatic\_dec.  
 DR InterPro; IPR02129; Pyridoxal\_dec.  
 DR Pfam; PF0282; Pyridoxal\_dec; 1.  
 DR PRINTS; PR00800; YHDCRBOXLASE.  
 DR PROSITE; PS00392; DDC GAD HDC YDC; 1.  
 KW Hypothetical protein; Lyase; Pyridoxal phosphate.  
 SQ SEQUENCE 705 AA; 79710 MW; BBBBDP83582FEC6 CRC64;

Query Match 51.8%; Score 57; DB 2; Length 705;  
 Best Local Similarity 41.2%; Pred. No. 1.2e+02;  
 Matches 7; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

QY 6 FEKXXXXXXWSHPQF 22  
 DB 136 FEKLIMPGITHMQHPRF 152

Search completed: March 2, 2005, 12:44:26  
 Job time : 91.1707 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 108.098 Seconds  
(without alignments)  
85.869 Million cell updates/sec

Title: SEQ12  
Perfect score: 110  
Sequence: 1 whnpqfekxxxxxxxxwhnpqfek 24

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16Dec04:\*

1: geneseqp19808:\*\n2: geneseqp19908:\*\n3: geneseqp20008:\*\n4: geneseqp20018:\*\n5: geneseqp20028:\*\n6: geneseqp20038:\*\n7: geneseqp20038a:\*\n8: geneseqp20048:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	24	ABP60362	Abp60362 Streptocavi
2	94.5	85.9	35	ABP60369	Abp60369 Streptocavi
3	94	85.5	36	ABP60370	Abp60370 Streptocavi
4	58.2	117	5	AAU97558	AAU97558 Synthetic
5	58.2	117	5	AAU97553	AAU97553 Synthetic
6	58.2	117	5	AAU97557	AAU97557 Synthetic
7	58.2	117	5	AAU97559	AAU97559 Synthetic
8	58.2	117	5	AAU97555	AAU97555 Synthetic
9	58.2	117	5	AAU97556	AAU97556 Synthetic
10	58.2	117	5	AAU97552	AAU97552 Synthetic
11	58.2	117	5	AAU97560	AAU97560 Synthetic
12	58.2	118	5	AAU97554	AAU97554 Synthetic
13	57.3	254	2	AAW93967	AAW93967 Plasmid p
14	57.3	254	2	AAW93969	AAW93969 Plasmid p
15	57.3	254	2	AAW93966	AAW93966 Plasmid p
16	57.3	396	4	AAW93966	AAW93966 Plasmid p
17	57.3	396	4	AAW93966	AAW93966 Plasmid p
18	57.3	396	4	AAW93966	AAW93966 Plasmid p
19	57.3	448	4	AAW93966	AAW93966 Plasmid p
20	57.3	448	4	AAW93966	AAW93966 Plasmid p
21	57.3	448	4	AAW93966	AAW93966 Plasmid p
22	57.3	448	4	AAW93966	AAW93966 Plasmid p
23	57.3	448	4	AAW93966	AAW93966 Plasmid p
24	57.3	448	4	AAW93966	AAW93966 Plasmid p
25	57.3	448	4	AAW93966	AAW93966 Plasmid p

26	61.5	55.9	697	8	ADS20227	Ads20227 Human agg
27	61	55.5	183	6	ADA27294	Ada27294 Plasmid p
28	61	55.5	183	6	ADA00703	Ada00703 Modified
29	61	55.5	400	6	ADA27291	Ada27291 Plasmid p
30	61	55.5	400	6	ADA27292	Ada27292 Plasmid p
31	61	55.5	400	6	ADA00700	Ada00700 Modified
32	61	55.5	400	6	ADA00707	Ada00707 Modified
33	61	55.5	400	6	ADA00701	Ada00701 Modified
34	61	55.5	467	8	ADP18461	Adp18461 Arthrobac
35	60	54.5	646	8	ADS20251	Ads20251 Human agg
36	60	54.5	661	6	AAE38369	Aae38369 Human agg
37	60	54.5	763	7	ADB85492	Adb85492 Human agg
38	60	54.5	772	8	ADP87577	Adp87577 Human agg
39	60	54.5	858	8	ADS20230	Ads20230 Human agg
40	59.5	53.6	633	8	ADS20255	Ads20255 Human agg
41	59	53.6	19	6	ABG74881	Abg74881 Bacteriop
42	59	53.6	19	6	ABG74882	Abg74882 Bacteriop
43	59	53.6	19	6	ABG74880	Abg74880 Bacteriop
44	59	53.6	19	8	ADI39157	Adi39157 Construct
45	59	53.6	19	8	ADI39155	Adi39155 Construct

#### ALIGNMENTS

RESULT 1	ABP60362	ABP60362 standard; peptide; 24 AA.
XX	XX	XX
AC	ABP60362;	
XX	XX	XX
DT	28-MAR-2003 (first entry)	
XX	XX	XX
DE	Streptavidin binding peptide SEQ ID NO 3.	
XX	XX	XX
KW	Streptavidin; protein chip; microtitre plate; detection.	
XX	XX	XX
OS	Synthetic.	
XX	XX	XX
FH	Key	Location/Qualifiers
FT	Misc-difference 9..16	
FT		/label= unknown
XX	XX	XX
PN	DE10113776-AL.	
XX	XX	XX
PD	02-OCT-2002.	
XX	XX	XX
PF	21-MAR-2001; 2001DE-01013776.	
XX	XX	XX
PR	21-MAR-2001; 2001DE-01013776.	
XX	XX	XX
PA	(BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.	
XX	XX	XX
PI	Schmidt T;	
XX	XX	XX
DR	WPI, 2003-031166/03.	
XX	XX	XX
PT	New isolated peptide, useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.	
PT		
PS	Disclosure; Page 4; 18pp; German.	
XX	XX	XX
CC	The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides stronger binding than a single tag, but are displaced by a competitor. (I) does not interfere with the function of attached proteins (II) (so it may not be essential to remove it); facilitates detection and has easily	



Query Match	Best Local Similarity	Score	DB	Length
Matches	9/	Conservative	8/	Mismatches
			0/	Indels
			0/	Gaps
				0/

```

QY      8 KXXXXXXXXXWSHPQEK 24
        |:::|||||
DB      .101 KGNVSCPSAMSHQPEK 117

RESULT 5
AAU97553
ID      AAU97553 standard; protein; 117 AA.
AC      AAU97553;
XX      13-AUG-2002 (first entry)
DT      13-AUG-2002 (first entry)
DB      Synthetic cadmium/mercury ion binding chelon protein #1.
KW      Mercuric ion; contaminated soil; ground water; hydroponic solution;
KW      irrigation water; waste stream; contaminated aqueous medium;
KW      biological fluid; gastrointestinal tract; chelon protein;
KW      enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;
KW      heavy metal binding protein.
XX      Synthetic.
OS      WO200230962-A2.
PN      18-APR-2002.
PD      12-OCT-2001; 2001WO-US031819.
PF      12-OCT-2000; 2000US-0240465P.
XX      12-OCT-2000; 2000US-0240465P.
PR      (UYGE-) UNIV GEORGIA RES FOUND INC.
XX      Sumners AO, Caguiat JI;
XX      WPI, 2002-435437/46.
XX      Novel non-naturally occurring recombinant DNA molecule encoding a chelon
PT      protein useful for binding divalent cation mercury from contaminated
PR      soil, water, aqueous medium including biological fluids.
XX      Claim 4; Page 22; 42pp; English.
XX      The present invention relates to a new non-naturally occurring
CC      recombinant DNA molecule comprising a sequence encoding a chelon protein
CC      which binds mercuric ions. The invention is useful for recombinantly
CC      producing a protein in a host-cell, by infecting or transforming a host
CC      cell capable of expressing a chelon coding sequence with a vector
CC      comprising a promoter active in the host cell operably linked to a coding
CC      region for the protein to produce a recombinant host cell and culturing
CC      the recombinant host cell under conditions, where DNA is expressed. The
CC      nucleic acid encoding the chelon protein is useful for binding divalent
CC      mercuric ions, to take up, sequester and concentrate the heavy metal ions
CC      from contaminated soil, ground water, hydroponic solutions or irrigation
CC      water of waste streams. The DNA of the invention, when immobilised onto a
CC      solid support, is useful for concentrating heavy metal ions from
CC      contaminated environment waste streams or contaminated aqueous medium
CC      including biological fluids. The nucleic acid, when recombinantly
CC      expressed in enteric bacteria (which are nontoxicogenic and nonpathogenic),
CC      is suitable for use in the in vivo sequestration and elimination of
CC      mercuric ion from gastrointestinal tracts of animals or humans exposed to
CC      toxic metal ions such as mercury and/or cadmium. The molecules of the
CC      invention are also useful in water treatment resins. The nucleic acid of
CC      the invention is highly specific and binds divalent cation such as
CC      mercury or cadmium with high affinity. The present amino acid sequence
CC      represents one of a collection (AAU97553-AAU97560) of synthetic
CC      cadmium/mercury ion binding chelon proteins of the invention. This
CC      sequence is one of the heavy metal binding proteins termed chelons of the
CC      invention
SQ      Sequence 117 AA;

```



CC sequence is one of the heavy metal binding proteins termed chelons of the  
 CC invention

XX Sequence 117 AA;

Query Match 58.2%; Score 64; DB 5; Length 117;

Best Local Similarity 52.9%; Pred. No. 3;

Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY 8 KXXXXXXXXXSHPOFEK 24

Db 101 KGNVCPSPASHPOFEK 117

RESULT 8

AAU97555

ID AAU97555 standard; protein; 117 AA.

XX AAU97555;

DT 13-AUG-2002 (first entry)

DE Synthetic cadmium/mercury ion binding chelon protein #3.

XX Mercuric ion; contaminated soil; ground water; hydroponic solution;

KW irrigation water; waste stream; contaminated aqueous medium;

KW biological fluid; gastrointestinal tract; chelon protein;

KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;

KW heavy metal binding protein.

OS Synthetic.

XX WO200230962-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US031819.

XX 12-OCT-2000; 2000US-0240465P.

PR (UYGE-) UNIV GEORGIA RES FOUND INC.

PA Sumners AO, Caguiat JJ;

PI WPI; 2002-435437/46.

DR Novel non-naturally occurring recombinant DNA molecule encoding a chelon  
 PT protein useful for binding divalent cation mercury from contaminated  
 PT soil, water, aqueous medium including biological fluids.

XX Claim 4; Page 22; 42pp; English.

CC The present invention relates to a new non-naturally occurring  
 CC recombinant DNA molecule comprising a sequence encoding a chelon protein  
 CC which binds mercuric ions. The invention is useful for recombinantly  
 CC producing a protein in a host-cell, by infecting or transforming a host  
 CC cell capable of expressing a chelon coding sequence with a vector  
 CC comprising a promoter active in the host cell operably linked to a coding  
 CC region for the protein to produce a recombinant host cell and culturing  
 CC the recombinant host cell under conditions, where DNA is expressed. The  
 CC nucleic acid encoding the chelon protein is useful for binding divalent  
 CC mercuric ions, to take up, sequester and concentrate the heavy metal ions  
 CC from contaminated soil, ground water, hydroponic solutions or irrigation  
 CC water of waste streams. The DNA of the invention, when immobilised onto a  
 CC solid support, is useful for concentrating heavy metal ions from  
 CC contaminated environment waste streams or contaminated aqueous medium  
 CC including biological fluids. The nucleic acid, when recombinantly  
 CC expressed in enteric bacteria (which are non-toxicogenic and nonpathogenic),  
 CC is suitable for use in the in vivo sequestration and elimination of  
 CC mercuric ion from gastrointestinal tracts of animals or humans exposed to  
 CC toxic metal ions such as mercury and/or cadmium. The molecules of the  
 CC invention are also useful in water treatment resins. The nucleic acid of  
 CC the invention is highly specific and binds divalent cation such as

CC mercury or cadmium with high affinity. The present amino acid sequence  
 CC represents one of a collection (AAU97553-AAU97560) of synthetic  
 CC cadmium/mercury ion binding chelon proteins of the invention. This  
 CC sequence is one of the heavy metal binding proteins termed chelons of the  
 CC invention

XX Sequence 117 AA;

Query Match 58.2%; Score 64; DB 5; Length 117;

Best Local Similarity 52.9%; Pred. No. 3;

Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY 8 KXXXXXXXXXSHPOFEK 24

Db 101 KGNVCPSPASHPOFEK 117

RESULT 9

AAU97556

ID AAU97556 standard; protein; 117 AA.

XX AAU97556;

DT 13-AUG-2002 (first entry)

DE Synthetic cadmium/mercury ion binding chelon protein #4.

XX Mercuric ion; contaminated soil; ground water; hydroponic solution;

KW irrigation water; waste stream; contaminated aqueous medium;

KW biological fluid; gastrointestinal tract; chelon protein;

KW enteric bacteria; toxic metal ion; mercury; cadmium; divalent cation;

KW heavy metal binding protein.

OS Synthetic.

XX WO200230962-A2.

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US031819.

XX 12-OCT-2000; 2000US-0240465P.

PR (UYGE-) UNIV GEORGIA RES FOUND INC.

PA Sumners AO, Caguiat JJ;

PI WPI; 2002-435437/46.

DR Novel non-naturally occurring recombinant DNA molecule encoding a chelon  
 PT protein useful for binding divalent cation mercury from contaminated  
 PT soil, water, aqueous medium including biological fluids.

XX Claim 4; Page 22; 42pp; English.

CC The present invention relates to a new non-naturally occurring  
 CC recombinant DNA molecule comprising a sequence encoding a chelon protein  
 CC which binds mercuric ions. The invention is useful for recombinantly  
 CC producing a protein in a host-cell, by infecting or transforming a host  
 CC cell capable of expressing a chelon coding sequence with a vector  
 CC comprising a promoter active in the host cell operably linked to a coding  
 CC region for the protein to produce a recombinant host cell and culturing  
 CC the recombinant host cell under conditions, where DNA is expressed. The  
 CC nucleic acid encoding the chelon protein is useful for binding divalent  
 CC mercuric ions, to take up, sequester and concentrate the heavy metal ions  
 CC from contaminated soil, ground water, hydroponic solutions or irrigation  
 CC water of waste streams. The DNA of the invention, when immobilised onto a  
 CC solid support, is useful for concentrating heavy metal ions from  
 CC contaminated environment waste streams or contaminated aqueous medium  
 CC including biological fluids. The nucleic acid, when recombinantly  
 CC expressed in enteric bacteria (which are non-toxicogenic and nonpathogenic),  
 CC is suitable for use in the in vivo sequestration and elimination of  
 CC mercuric ion from gastrointestinal tracts of animals or humans exposed to

CC contaminated environment waste streams or contaminated aqueous medium

CC from contaminated soil, ground water, hydroponic solutions or irrigation

Query Match	58.2%	Score 64	DB 5	Length 117
Best Local Similarity	52.9%	Pred. NO. 3		
Matches 9, Conservative		8, Mismatches	0, Indels	0, Gaps

```
Qy      8 |XXXXXXXXXXSHPOFEK 24
         |:.....| | | |
Db      101 |GNVSCPSAMSHPOFEK 117
```

RESULT 12  
AAU97554  
ID AAU97554 standard; protein; 118 AA.

DT 13-AUG-2002, (first entry)

**Synthetic cadmium/mercury ion binding chelon protein #2.**

KW Mercuric ion; contaminated soil; ground water; hydroponic solution;

KW biological fluid; gastrointestinal tract; chelon protein;

KW heavy metal binding protein. ....

OS Synthetic.

PN WO200230962-A2

PD 18-APR-2002.

PF 12-OCT-2001; 2001WO-US031819

PR 12-OCT-2000; 2000US-0240465P.

PA (UYGE-) UNIV GEORGIA RES FOUND INC.

PI Summers AO, Cagulat JJ;

DR WPI; 2002-435437/46.

PT Novel non-naturally occurring recombinant DNA molecule encoding a chelon  
PT protein useful for binding divalent cation mercury from contaminated  
PT soil, water, aqueous medium including biological fluids.

PS Claim 4; Page 22; 42pp; English.

CC The present invention relates to a new non-naturally occurring  
CC recombinant DNA molecule comprising a sequence encoding a chelon protein  
CC which binds mercuric ions. The invention is useful for recombinantly  
CC producing a protein in a host-cell, by infecting or transforming a host  
CC cell capable of expressing a chelon coding sequence with a vector  
CC comprising a promoter active in the host cell operably linked to a coding  
CC region for the protein to produce a recombinant host cell and culturing  
CC the recombinant host cell under conditions, where DNA is expressed. The

CC nucleic acid encoding the chelon protein is useful for binding divalent  
CC mercuric ions, to take up, sequester and concentrate the heavy metal ions  
CC from contaminated soil, ground water, hydroponic solutions or irrigation  
CC water of waste streams. The DNA of the invention, when immobilised onto a  
CC solid support, is useful for concentrating heavy metal ions from  
CC contaminated environment waste streams or contaminated aqueous medium  
CC including biological fluids. The nucleic acid, when recombinantly  
CC expressed in enteric bacteria (which are nontoxic and nonpathogenic),  
CC is suitable for use in the in vivo sequestration and elimination of  
CC mercuric ion from gastrointestinal tracts of animals or humans exposed to  
CC toxic metal ions such as mercury and/or cadmium. The molecules of the  
CC invention are also useful in water treatment resins. The nucleic acid of  
CC the invention is highly specific and binds divalent cation such as  
CC mercury or cadmium with high affinity. The present amino acid sequence  
CC represents one of a collection (AAU07553-AAU07560) of synthetic  
CC cadmium/mercury ion binding chelon proteins of the invention. This  
CC sequence is one of the heavy metal binding proteins termed chelons of the  
CC invention

**SQ** Sequence 118 AA;

Query Match	58.2%	Score 64	DB 5	Length 118
Best Local Similarity	52.9%	Pred. No. 3		
Matches 9	Conservative 8	Mismatches 0	Indels 0	Gaps 0

```
QY      8 KXXXXXXXXXWHPQFEK 24  
         |:::~::~|||~|||  
Db     102 KGNVSCPSAWSHPQFEK 116
```

RESULT 13	
AAW93967	
ID	AAW93967 standard; protein; 205 AA.

DT 02-JUL-1999 (first entry)

DE Plasmid pBBP21 protein fragment.

KW Lipocalin; anticalin; peptide loop; leaflet structure; antibody; therapy;

KM cumour imaging; bilin binding protein; DsbC protein

OS Synthetic

FH	Key	Location/Qualifiers
1	1	1
2	2	2
3	3	3
4	4	4
5	5	5
6	6	6
7	7	7
8	8	8
9	9	9
10	10	10
11	11	11
12	12	12
13	13	13
14	14	14
15	15	15
16	16	16
17	17	17
18	18	18
19	19	19
20	20	20
21	21	21
22	22	22
23	23	23
24	24	24
25	25	25
26	26	26
27	27	27
28	28	28
29	29	29
30	30	30
31	31	31
32	32	32
33	33	33
34	34	34
35	35	35
36	36	36
37	37	37
38	38	38
39	39	39
40	40	40
41	41	41
42	42	42
43	43	43
44	44	44
45	45	45
46	46	46
47	47	47
48	48	48
49	49	49
50	50	50
51	51	51
52	52	52
53	53	53
54	54	54
55	55	55
56	56	56
57	57	57
58	58	58
59	59	59
60	60	60
61	61	61
62	62	62
63	63	63
64	64	64
65	65	65
66	66	66
67	67	67
68	68	68
69	69	69
70	70	70
71	71	71
72	72	72
73	73	73
74	74	74
75	75	75
76	76	76
77	77	77
78	78	78
79	79	79
80	80	80
81	81	81
82	82	82
83	83	83
84	84	84
85	85	85
86	86	86
87	87	87
88	88	88
89	89	89
90	90	90
91	91	91
92	92	92
93	93	93
94	94	94
95	95	95
96	96	96
97	97	97
98	98	98
99	99	99
100	100	100

FT	/label= signal_peptide
1	1
2	1
3	1
4	1
5	1
6	1
7	1
8	1
9	1
10	1
11	1
12	1
13	1
14	1
15	1
16	1
17	1
18	1
19	1
20	1
21	1
22	1
23	1
24	1
25	1
26	1
27	1
28	1
29	1
30	1
31	1
32	1
33	1
34	1
35	1
36	1
37	1
38	1
39	1
40	1
41	1
42	1
43	1
44	1
45	1
46	1
47	1
48	1
49	1
50	1
51	1
52	1
53	1
54	1
55	1
56	1
57	1
58	1
59	1
60	1
61	1
62	1
63	1
64	1
65	1
66	1
67	1
68	1
69	1
70	1
71	1
72	1
73	1
74	1
75	1
76	1
77	1
78	1
79	1
80	1
81	1
82	1
83	1
84	1
85	1
86	1
87	1
88	1
89	1
90	1
91	1
92	1
93	1
94	1
95	1
96	1
97	1
98	1
99	1
100	1

FT /label= fusion\_peptide

FT . binding protein and a strep-Tag II region."

PN WO9916873-A1.

PD 08-APR-1999

PF 25-SEP-1998; 98WO-DE002898

PR 26-SEP-1997; 97DE-01042706

PA (SKER/) SKERRA A.

PI Skerra A, Beste G, Schmidt F, Stibora T, ...

DR WPI; 1999-255093/21

PT New anticalins produced by mutation of lipocalin family protein.

PS Disclosure; Page 70-72; 80pp; German.









*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 74.6341 Seconds  
(without alignments)  
105.489 Million cell updates/sec

Title: SEQ12  
Perfect score: 110  
Sequence: 1 whpghfexkxxxxxxwhpghfek 24

Scoring table: BLOSUM62GX  
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Listing first 45 summaries

Database : Published Applications AA:\*

1:	/cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/2/pubppaa/PCT_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/2/pubppaa/PCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/2/pubppaa/US09B_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/2/pubppaa/US09C_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep.*
13:	/cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/2/pubppaa/US10B_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/2/pubppaa/US10C_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/2/pubppaa/US10D_PUBCOMB.pep.*
17:	/cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep.*
18:	/cgn2_6/ptodata/2/pubppaa/US11_NEW_PUB.pep.*
19:	/cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep.*
20:	/cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	110	100.0	24	US-10-026-578B-3	Sequence 3, Appli
2	94	85.5	36	US-10-026-578B-10	Sequence 10, Appli
3	81	73.6	36	US-10-026-578B-11	Sequence 11, Appli
4	81	58.2	21	US-09-809-517A-30	Sequence 30, Appli
5	64	58.2	22	US-09-809-517A-33	Sequence 33, Appli
6	58.2	117	10	US-09-977-137A-4	Sequence 4, Appli
7	58.2	117	10	US-09-977-137A-5	Sequence 5, Appli
8	58.2	117	10	US-09-977-137A-7	Sequence 7, Appli
9	58.2	117	10	US-09-977-137A-8	Sequence 8, Appli
10	58.2	117	10	US-09-977-137A-9	Sequence 9, Appli
11	58.2	117	10	US-09-977-137A-10	Sequence 10, Appli
12	58.2	117	10	US-09-977-137A-11	Sequence 11, Appli
13	58.2	117	10	US-09-977-137A-12	Sequence 12, Appli

14	64	58.2	118	10	US-09-977-137A-6	Sequence 6, Appl
15	62	56.4	845	16	US-10-628-432-40	Sequence 40, Appl
16	61.5	55.9	485	16	US-10-628-432-47	Sequence 47, Appl
17	61.5	55.9	697	16	US-10-628-432-24	Sequence 24, Appl
18	61	55.5	24	9	US-09-809-517A-31	Sequence 31, Appl
19	61	55.5	25	9	US-09-809-517A-34	Sequence 34, Appl
20	61	55.5	245	17	US-10-887-228A-1	Sequence 1, Appl
21	61	55.5	246	17	US-10-887-228A-9	Sequence 9, Appl
22	61	55.5	252	17	US-10-887-228A-5	Sequence 5, Appl
23	60	54.5	646	16	US-10-628-432-49	Sequence 49, Appl
24	60	54.5	661	15	US-10-354-983-8	Sequence 8, Appl
25	60	54.5	763	15	US-10-358-283-15	Sequence 15, Appl
26	60	54.5	858	16	US-10-628-432-27	Sequence 27, Appl
27	59.5	54.1	633	16	US-10-628-432-53	Sequence 53, Appl
28	59	53.6	24	14	US-10-026-578B-4	Sequence 4, Appl
29	59	53.6	126	9	US-09-815-242-12508	Sequence 12508, A
30	59	53.6	231	9	US-09-815-242-5760	Sequence 5760, Ap
31	59	53.6	298	15	US-10-282-122A-44219	Sequence 44219, A
32	57	51.8	650	15	US-10-369-493-5480	Sequence 5480, Ap
33	57	51.8	847	15	US-10-389-566-832	Sequence 832, App
34	57	51.8	1059	15	US-10-389-566-834	Sequence 834, App
35	57	51.8	1069	15	US-10-389-566-833	Sequence 833, App
36	57	51.8	1069	16	US-10-437-963-143424	Sequence 143424, A
37	56.5	51.4	134	16	US-10-437-963-104065	Sequence 104065, A
38	56.5	51.4	252	16	US-10-437-963-200150	Sequence 200150, A
39	56.5	51.4	629	15	US-10-282-122A-77703	Sequence 77703, A
40	55.5	50.5	158	15	US-10-424-599-184769	Sequence 184769, A
41	55.5	50.5	162	14	US-10-204-887-94	Sequence 94, Appl
42	55.5	50.5	212	15	US-10-424-599-184075	Sequence 184075, A
43	55	50.0	56	11	US-09-833-245-1103	Sequence 1103, Ap
44	55	50.0	56	11	US-09-833-245-1105	Sequence 1105, Ap
45	55	50.0	329	15	US-10-449-831A-204	Sequence 204, App

## ALIGNMENTS

RESULT 1  
US-10-026-578B-3  
Sequence 3, Application US/10026578B  
Publication No. US20030083474A1  
GENERAL INFORMATION:  
APPLICANT: IBA (GmbH)  
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta  
FILE REFERENCE: 100810.01US1  
CURRENT APPLICATION NUMBER: US/10/026,578B  
CURRENT FILING DATE: 2002-11-11  
PRIOR APPLICATION NUMBER: DE 101 13 776.1  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: PCT/EP01/11846  
PRIOR FILING DATE: 2001-10-12  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
LENGTH: 24  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Synthetic Peptide  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (9)..(9)  
OTHER INFORMATION: X represents a single amino acid at the position indicated  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (10)..(10)  
OTHER INFORMATION: X represents a single amino acid at the position indicated  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (11)..(11)  
OTHER INFORMATION: X represents a single amino acid at the position indicated  
FEATURE:



Matches	9,	Conservative	8,	Mismatches	0,	Indels	0,	Gaps	0
QY	8	XXXXXXXXXX	MSHPREK	24					
		: : : : :	: : :						
Db	5	KDDDDKGA	MSHPREK	21					

RESULT 5  
US-09-809-517A-33

```

Sequence 33 Application US/09809517A
Patent No. US20020034733A1
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly) peptides/proteins
TITLE OF INVENTION: particles via disulfide bonds
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
PRIORITY FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIORITY FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIORITY FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.0
SEQ ID NO 33
LENGTH: 22
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33

```

```

Query Match      58.2% Score 64 DB 9 Length 22;
Best Local Similarity 52.9% Pred No. 0.52;
Matches          9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;
QY              8 KXXXXXXXXXSHPOPEK 24
                | : : : : : : : : : :
Db               6 KDDDDKGAPWSHPDPEK 22

```

RESULT 6  
 US-09-977-137A-4  
 Sequence 4, Application US/09977137A  
 Publication NO. US20030104524A1  
 GENERAL INFORMATION:  
 APPLICANT: Summers, Anne O.  
 APPLICANT: Caguiat, Jonathan  
 TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and  
 TITLE OF INVENTION: Methods  
 FILE REFERENCE: 79-00  
 CURRENT APPLICATION NUMBER: US/09/977,137A  
 CURRENT FILING DATE: 2001-10-12  
 PRIOR APPLICATION NUMBER: US 60/240,465  
 PRIOR FILING DATE: 2000-10-12  
 NUMBER OF SEQ ID NOS: 18  
 SOFTWARE: PatentIn Ver. 2.0  
 SEQ ID NO 4  
 LENGTH: 117  
 TYPE: PRT  
 ORGANISM: Artificial Sequence  
 FEATURE:  
 OTHER INFORMATION: Description of Artificial Sequence: chelon  
 US-09-977-137A-4

Query Match	58.2%	Score 64	DB 10	Length 117
Best Local	Similarity 52.9%	Pred No. 3.7		
Matches	9	Conservative	8	Mismatches 0; Indels 0; Gaps 0
Qy	8	KXXXXXXXXXSHPOFEK	24	
Qb	101	KGNNSCPSAMSHPOFEK	117	

RESULT 7:  
US-09-977-137A-5  
; Sequence 5, Application US/09977137A  
; Publication No. US20030104524A1

```

1  TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
2  TITLE OF INVENTION: Methods
3  FILE REFERENCE: 79-00
4  CURRENT APPLICATION NUMBER: US/09/977,137A
5  CURRENT FILING DATE: 2001-10-12
6  PRIOR APPLICATION NUMBER: US 60/240,465
7  PRIOR FILING DATE: 2000-10-12
8  NUMBER OF SEQ ID NOS: 18
9  SOFTWARE: PatentIn Ver. 2.0
10 SEQ ID NO 5
11
12 LENGTH: 117
13
14 TYPE: PRT
15
16 ORGANISM: Artificial Sequence
17 FEATURE:
18
19 OTHER INFORMATION: Description of Artificial Sequence: chelon
20 US-09-977-137A-5

```

```

Query Match          58.2%   Score 64; DB 10; Length 117;
      Best Local Similarity    52.9%   Pred. No. 3.7'
Matches      9; Conservative      8; Mismatches      0; Indels      0; Gaps      0;

QY      8 KXXXXXXXXXSHPOFEK 24
        |:::|:::|:::|:::|
Db       101 KGNVSCPEAWSHPOFEK 117

```

```

RESULT 8
US-09-977-137A-7
: Sequence 7, Application US/09977137A
: Publication NO. US20030104524A1
: GENERAL INFORMATION:
: APPLICANT: Summers, Anne O.
: APPLICANT: Caguiat, Jonathan
: TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
: TITLE OF INVENTION: Methods
: FILE REFERENCE: 79-00
: CURRENT APPLICATION NUMBER: US/09/977,137A
: CURRENT FILING DATE: 2001-10-12
: PRIOR APPLICATION NUMBER: US 60/240,465
: PRIOR FILING DATE: 2000-10-12
: NUMBER OF SEQ ID NOS: 18
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 7
: LENGTH: 117
: TYPE: PRT
: ORGANISM: Artificial Sequence
FEATURE:
: OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-7

```

Query Match	58.2%	Score 64	DB 10	Length 117
Best Local Similarity	52.9%	Pred. NO.	3.7	
Matches	9	Conservative	8	Mismatches 0; Indels 0; Gaps 0
QY	8 KXXXXXXXXXSHPOFEK	24		
DB	101 KGNVSCPEAWSHPOFEK	117		

RESULT 9  
US-09-977-137A-8  
; Sequence 8, Application US/09977137B  
; Publication No. US2003010452A1  
; GENERAL INFORMATION:  
; APPLICANT: Summers, Anne O.

```

; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 8
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-8

Query Match          58.2%; Score 64; DB 10; Length 117;
Best Local Similarity 52.9%; Pred. No. 3.7;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXWSHPQFEK 24
        |:::|||||
DB      101 KGNVSCPSAMSHPOFEK 117

RESULT 10
US-09-977-137A-9
; Sequence 9; Application US/09977137A
; Publication No. US20030104524A1
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 9
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-9

Query Match          58.2%; Score 64; DB 10; Length 117;
Best Local Similarity 52.9%; Pred. No. 3.7;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXWSHPQFEK 24
        |:::|||||
DB      101 KGNVSCPSAMSHPOFEK 117

RESULT 11
US-09-977-137A-10
; Sequence 10; Application US/09977137A
; Publication No. US20030104524A1
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-10
```

```

; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-10

Query Match          58.2%; Score 64; DB 10; Length 117;
Best Local Similarity 52.9%; Pred. No. 3.7;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXWSHPQFEK 24
        |:::|||||
DB      101 KGNVSCPSAMSHPOFEK 117

RESULT 12
US-09-977-137A-11
; Sequence 11; Application US/09977137A
; Publication No. US20030104524A1
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-11

Query Match          58.2%; Score 64; DB 10; Length 117;
Best Local Similarity 52.9%; Pred. No. 3.7;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXWSHPQFEK 24
        |:::|||||
DB      101 KGNVPCPSAMSHPOFEK 117

RESULT 13
US-09-977-137A-12
; Sequence 12; Application US/09977137A
; Publication No. US20030104524A1
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
US-09-977-137A-12
```



*This Page Blank (uspio)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 27.3171 Seconds  
(without alignments)  
65.585 Million cell updates/sec

Title: SEQ12

Perfect score: 110

Sequence: 1 wshpqfexkxxxxxxxxxwshpqfex 24

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

Issued Patents AA: \*  
1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep: \*  
2: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep: \*  
3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep: \*  
4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep: \*  
5: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep: \*  
6: /cgn2\_6/ptodata/1/1aa/backfile1.pep: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	58.2	21	4	US-09-809-517A-30 Sequence 30, Appl
2	64	58.2	22	4	US-09-809-517A-33 Sequence 33, Appl
3	64	58.2	117	4	US-09-977-137A-4 Sequence 4, Appl
4	64	58.2	117	4	US-09-977-137A-5 Sequence 5, Appl
5	64	58.2	117	4	US-09-977-137A-7 Sequence 7, Appl
6	64	58.2	117	4	US-09-977-137A-8 Sequence 8, Appl
7	64	58.2	117	4	US-09-977-137A-9 Sequence 9, Appl
8	64	58.2	117	4	US-09-977-137A-10 Sequence 10, Appl
9	64	58.2	117	4	US-09-977-137A-11 Sequence 11, Appl
10	64	58.2	117	4	US-09-977-137A-12 Sequence 12, Appl
11	64	58.2	118	4	US-09-977-137A-6 Sequence 6, Appl
12	61	55.5	24	4	US-09-809-517A-31 Sequence 31, Appl
13	61	55.5	25	4	US-09-809-517A-34 Sequence 34, Appl
14	55	50.0	699	4	US-09-538-092-995 Sequence 995, App
15	54	49.1	83	4	US-09-732-210-1454 Sequence 1454, Ap
16	54	49.1	512	3	US-09-413-814-8 Sequence 8, Appl
17	54	49.1	521	3	US-08-504-048-9 Sequence 9, Appl
18	53	48.2	10	4	US-09-809-517A-6 Sequence 6, Appl
19	53	48.2	233	4	US-09-328-352-5498 Sequence 5498, Ap
20	52	47.3	2324	4	US-09-902-540-9732 Sequence 9732, Ap
21	51	46.4	8	3	US-08-948-097-2 Sequence 2, Appl
22	51	46.4	8	3	US-09-382-950-7 Sequence 7, Appl
23	51	46.4	8	3	US-09-382-7358-8 Sequence 8, Appl
24	51	46.4	8	4	US-09-619-103-9 Sequence 9, Appl
25	51	46.4	8	4	US-10-104-218-5 Sequence 5, Appl
26	51	46.4	8	4	US-09-809-517A-9 Sequence 9, Appl
27	51	46.4	86	4	US-09-198-452A-677 Sequence 677, App

28	51	46.4	94	4	US-09-438-185A-640 Sequence 640, App
29	51	46.4	108	4	US-09-023-905A-32 Sequence 32, Appl
30	51	46.4	204	4	US-09-248-796A-16426 Sequence 16426, A
31	51	46.4	406	4	US-09-248-796A-16826 Sequence 16826, A
32	51	46.4	420	4	US-09-949-016-9364 Sequence 9364, Ap
33	51	46.4	476	4	US-09-949-016-10314 Sequence 10314, A
34	51	46.4	537	4	US-09-949-016-5954 Sequence 5954, Ap
35	50	45.5	158	4	US-09-621-976-4227 Sequence 4227, Ap
36	50	45.5	273	4	US-09-328-352-6417 Sequence 6417, Ap
37	50	45.5	356	4	US-09-198-452A-526 Sequence 526, App
38	50	45.5	487	4	US-09-602-777A-420 Sequence 420, App
39	50	45.5	527	4	US-09-600-985-1 Sequence 1, Appl
40	50	45.5	527	4	US-09-600-985-2 Sequence 2, Appl
41	50	45.5	527	4	US-09-600-985-3 Sequence 3, Appl
42	50	45.5	574	4	US-09-438-185A-489 Sequence 489, App
43	50	45.5	645	4	US-09-547-435-26 Sequence 26, Appl
44	50	45.5	771	4	US-09-547-435-28 Sequence 28, Appl
45	50	45.5	867	4	US-09-547-435-24 Sequence 24, Appl

#### ALIGNMENTS

```
RESULT 1
US-09-809-517A-30
; Sequence 30, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on t
; TITLE OF INVENTION: Particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 21
; TYPE: prt
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30
Query Match          58.2%; Score 64; DB 4; Length 21;
Best Local Similarity 52.9%; Pred. No. 0.026;
Matches          9; Conservative      8; Mismatches      0; Indels      0; Gaps      0;
Oy      8 KXXXXXXXXXSHPOFEK 24
Db      5 KDDDDKGAFWSHPOFEK 21
US-09-809-517A-33
; Sequence 33, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; TITLE OF INVENTION: Particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
```

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 22
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33

Query Match          58.2%; Score 64; DB 4; Length 22;
Best Local Similarity 52.9%; Pred. No. 0.028;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy      8 KXXXXXXXXXWHPQFEK 24
Db      6 KDDDDKGAFWSHPOFEK 22

RESULT 3
US-09-977-137A-4
; Sequence 4, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4

Query Match          58.2%; Score 64; DB 4; Length 117;
Best Local Similarity 52.9%; Pred. No. 0.28;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy      8 KXXXXXXXXXWHPQFEK 24
Db      101 KGNVSCPSAWSHPOFEK 117

RESULT 4
US-09-977-137A-5
; Sequence 5, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
```

```

US-09-977-137A-5

Query Match          58.2%; Score 64; DB 4; Length 117;
Best Local Similarity 52.9%; Pred. No. 0.28;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy      8 KXXXXXXXXXWHPQFEK 24
Db      101 KGNVSCPSAWSHPOFEK 117

RESULT 5
US-09-977-137A-7
; Sequence 7, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 7
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-7

Query Match          58.2%; Score 64; DB 4; Length 117;
Best Local Similarity 52.9%; Pred. No. 0.28;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy      8 KXXXXXXXXXWHPQFEK 24
Db      101 KGNVSCPSAWSHPOFEK 117

RESULT 6
US-09-977-137A-8
; Sequence 8, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-8

Query Match          58.2%; Score 64; DB 4; Length 117;
Best Local Similarity 52.9%; Pred. No. 0.28;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

Qy      8 KXXXXXXXXXWHPQFEK 24
```

```
Db      101 KGNVSCPSAMSHPOFEK 117
|:|||||
RESULT 7
US-09-977-137A-9
; Sequence 9, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 9
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-9

Query Match      58.2%; Score 64; DB 4; Length 117;
Best Local Similarity 52.9%; Pred. No. 0.28;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXXXXSHPOFEK 24
|:|||||
Db      101 KGNVSCPSAMSHPOFEK 117

RESULT 8
US-09-977-137A-10
; Sequence 10, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 10
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-10

Query Match      58.2%; Score 64; DB 4; Length 117;
Best Local Similarity 52.9%; Pred. No. 0.28;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXXXXSHPOFEK 24
|:|||||
Db      101 KGNVSCPSAMSHPOFEK 117

RESULT 9
US-09-977-137A-11
; Sequence 11, Application US/09977137A
```

```
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-11

Query Match      58.2%; Score 64; DB 4; Length 117;
Best Local Similarity 52.9%; Pred. No. 0.28;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXXXXSHPOFEK 24
|:|||||
Db      101 KGNVSCPSAMSHPOFEK 117

RESULT 10
US-09-977-137A-12
; Sequence 12, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-12

Query Match      58.2%; Score 64; DB 4; Length 117;
Best Local Similarity 52.9%; Pred. No. 0.28;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXXXXSHPOFEK 24
|:|||||
Db      101 KGNVSCPSAMSHPOFEK 117

RESULT 11
US-09-977-137A-6
; Sequence 6, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
```

```

; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 118
; TYPE: PRF
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-6

Query Match          58.2%; Score 64; DB 4; Length 118;
Best Local Similarity 52.9%; Pred. No. 0.29;
Matches 9; Conservative 8; Mismatches 0; Indels 0; Gaps 0;

QY 8 KXXXXXXXXXWHPQFEK 24
Db 102 KGNVSCPSAMSHPOFEK 118

RESULT 12
US-09-809-517A-31
; Sequence 31, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly) peptides/proteins on h
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
; LENGTH: 24
; TYPE: PRF
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-31

Query Match          55.5%; Score 61; DB 4; Length 24;
Best Local Similarity 41.7%; Pred. No. 0.086;
Matches 10; Conservative 10; Mismatches 0; Indels 4; Gaps 1;

QY 5 OPEK---XXXXXXXWHPQFEK 24
Db 1 EPEQKLISEEDINGAPWHPQFEK 24

RESULT 13
US-09-809-517A-34
; Sequence 34, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on h
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
```

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRF
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34

Query Match          55.5%; Score 61; DB 4; Length 25;
Best Local Similarity 41.7%; Pred. No. 0.091;
Matches 10; Conservative 10; Mismatches 0; Indels 4; Gaps 1;

QY 5 OPEK---XXXXXXXWHPQFEK 24
Db 2 EPEQKLISEEDINGAPWHPQFEK 25

RESULT 14
US-09-538-092-995
; Sequence 995, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Giot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; CURRENT FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuraPatSeqFormatter Version 0.9
; SEQ ID NO 995
; LENGTH: 699
; TYPE: PRF
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)...(0)
; OTHER INFORMATION: Polypeptide Accession Number P23327
US-09-538-092-995

Query Match          50.0%; Score 55; DB 4; Length 699;
Best Local Similarity 36.8%; Pred. No. 67;
Matches 7; Conservative 9; Mismatches 3; Indels 0; Gaps 0;

QY 1 WHPQFEKXXXXXXXXXWHP 19
Db 83 WHPQFEKEDVSKSEYGH 101

RESULT 15
US-09-732-210-1454
; Sequence 1454, Application US/09732210
; Patent No. 6573361
; GENERAL INFORMATION:
; APPLICANT: Bunkers, Greg J.
; APPLICANT: Liang, Jihong
; APPLICANT: Mitcanck, Cindy A.
; APPLICANT: Seale, Jeffrey W.
; APPLICANT: Wu, Yonnie S.
; TITLE OF INVENTION: Anti-fungal Proteins and Methods for Their Use
; FILE REFERENCE: 38-21(15036)B
; CURRENT APPLICATION NUMBER: US/09/732,210
; CURRENT FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/169,513
; PRIOR FILING DATE: 1999-12-07
; PRIOR APPLICATION NUMBER: US 60/169,340
; PRIOR FILING DATE: 1999-12-07
; NUMBER OF SEQ ID NOS: 1753
```

```

; SEQ ID NO 1454
;
; LENGTH: 83
;
; TYPE: prt
;
; ORGANISM: Chlamydia trachomatis
;
US-09-732-210-1454

```

Query Match 49.1%; Score 54; DB 4; Length 83;  
Best Local Similarity 26.1%; Pred. NO. 4.9;  
Matches 6; Conservative 14; Mismatches 3; Indels 0; Gaps 0;

```

QY      1 WSHPQFEKXXXXXXXXXWSHPQFE 23
          ||||:|:::||||:|:
Db      33 YSHPQYAKVNRDSSKYVAHNELD 55

```

Search completed: March 2, 2005, 12:25:37  
Job time : 28.3171 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 19.1057 Seconds  
(without alignments)  
125.901 Million cell updates/sec

Title: SEQ13  
Perfect score: 111  
Sequence: 1 wshpqfexkxxxxxxwshpqfex 25

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	58	52.3	1113	2 T20004	hypothetical prote
2	57	51.4	73	2 S28978	srp protein - phag
3	56	50.5	759	2 D81657	general secretion
4	55	49.5	103	2 E90973	hypothetical prote
5	55	49.5	103	2 A85821	unknown protein en
6	55	49.5	103	2 C85713	unknown protein en
7	55	49.5	103	2 A90904	hypothetical prote
8	55	49.5	315	2 T23597	hypothetical prote
9	55	49.5	318	2 S76226	hypothetical prote
10	55	49.5	330	2 AC2071	hypothetical prote
11	55	49.5	379	2 T23592	hypothetical prote
12	54	48.6	260	2 AB2080	hypothetical prote
13	54	48.6	308	2 G86464	P12612.10 protein
14	54	48.6	414	2 G84311	hypothetical prote
15	54	48.6	625	2 G72420	oligopeptidase ABC
16	54	48.6	958	2 S62413	probable sulfatase
17	54	48.6	2150	1 S27802	zinc finger protei
18	54	48.6	2150	2 T19450	hypothetical prote
19	53	47.7	83	1 C42645	ribosomal protein
20	53	47.7	83	2 F81664	ribosomal protein
21	53	47.7	437	2 A31142	gelatin, ovarian
22	52.5	47.3	220	1 S22516	S-allele-associated
23	52.5	47.3	220	2 A54660	histidine rich cal
24	52	46.8	108	2 P95385	protein (imported
25	52	46.8	226	2 E83081	conserved hypotet
26	52	46.8	323	2 E88642	protein C54E4.5 (l
27	52	46.8	434	2 G95297	probable hydrolase
28	52	46.8	451	2 G01227	recepin - human
29	52	46.8	633	2 D72284	oligopeptide ABC

30	52	46.8	760	2 E71499	probable gen. secr
31	52	46.8	957	2 T10633	hypothetical prote
32	51.3	46.4	71	1 IHER1	high potential pro
33	51	45.9	109	2 E89873	cysteine proteina
34	51	45.9	225	2 A75425	hypothetical prote
35	51	45.9	318	2 B75382	conserved hypotet
36	51	45.9	656	2 T01573	earl protein - mal
37	51	45.9	664	2 A47012	phenol 2-monooxyge
38	51	45.9	754	2 E86592	general secretion
39	51	45.9	754	2 D72032	general secretion
40	51	45.9	1003	2 T28654	transposase - Paen
41	51	45.9	3947	2 T52486	ferrichrome sidero
42	50.5	45.5	1078	2 S77162	DNA topoisomerase
43	50	45.0	86	2 C86570	S17 ribosomal prot
44	50	45.0	86	2 H72054	ribosomal protein
45	50	45.0	161	2 B42627	cement precursor p

## ALIGNMENTS

RESULT 1  
T20004  
hypothetical protein C47E12.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T20004  
R:Colles, L.  
submitted to the EMBL Data Library, January 1996  
A:Reference number: Z19210  
A:Accession: T20004  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1113 <WILL>  
A:Cross-references: UNIPROT:Q27481; EMBL:Z68882; P1DN:CAA93101.1; GSPDB:GN00022; CESP:C  
A:Experimental source: clone C47E12  
C:Genetics:  
A:Gene: CESP:C47E12.5  
A:Map position: 4  
A:Introns: 65/3; 112/2; 213/3; 574/3; 968/3; 1070/2  
C:Superfamily: ubiquitin-activating enzyme E1

Query Match  
Query Length 1113; Score 58; DB 2; Length 1113;  
Best Local Similarity 34.8%; Pred. No. 37;  
Matches 8; Conservative 11; Mismatches 4; Indels 0; Gaps 0;  
QY 2 SHPOFEKXXXXXXXXXWSHPOFE 24  
DB 355 SHVPEKSLVEBPFGIMDYAKFE 377

RESULT 2  
S28978  
srp protein - phage HK022  
C:Species: phage HK022  
C:Date: 07-Apr-1994 #sequence\_revision 07-Apr-1994 #text\_change 09-Jul-2004  
C:Accession: S28978; S16858  
R:Atkinson, B.L.; Gottesman, M.E.  
J. Mol. Biol. 227, 29-37, 1992  
A:Title: The Escherichia coli ropB60 mutation blocks antitermination by coliphage HK022  
A:Reference number: S28977; MUID:92595669; PMID:1522593  
A:Accession: S28978  
A:Molecule type: DNA  
A:Residues: 1-73 <ATK>  
C:Cross-references: UNIPROT:Q38575; EMBL:X60308; NID:G15753; P1DN:CAA42853.1; P1D:957915  
A:Genetics:  
A:Gene: srp  
A:Start codon: GTG

Query Match  
Best Local Similarity 51.4%; Score 57; DB 2; Length 73;  
Matches 8; Conservative 11; Mismatches 3; Indels 0; Gaps 0;







	48.6%;	Score 54;	DB 2;	Length 414;
Query Match	Similarity	32.0%;	Pred. NO. 43;	
Best Local	Conservative	11;	Mismatches	Gaps
Matches				
Oy	1 WSHPOFEKXXXXXXXXXSHPOFEK	25		
	: : : :			
Db	124 FSLSNFTKSTLRSCKDKDMSPEFEK	148		
	: : : :			

Search completed: March 2, 2005, 12:28:59  
Job time : 20.1057 secs

```
QY      1 WSHPQFEKXXXXXXXXXXWSHPQ 22
         |||:|:|:|:|:|:|:|:|:|
Db      535 WSDPRLDKIIEELKNTPWGNTQ 556
```



```
RC STRAIN=Friedlin;
RX MEDLINE=98146435; PubMed=9477341;
RA Ivens A.C., Lewis S.M., Bagnerzaden A., Zhang L., Chan H.M.,
RA Smith D.F.;
RT "A physical map of the Leishmania major Friedlin genome.";
RL Genome Res. 8:135-145(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Friedlin;
RA Clariom L., Tosato V., Bruschi C.V., Ivens A.C., Quail M.,
RA Rajandream M.A., Barrell B.G.;
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AL512295; CAC24652.1; -.
KW Hypothetical protein.
SQ SEQUENCE 270 AA; 29661 MW; 60829CD354CFB36 CRC64;

Query Match          56.8%; Score 63; DB 2; Length 270;
Best Local Similarity 34.8%; Pred. No. 9.8;
Matches 8; Conservative 13; Mismatches 2; Indels 0; Gaps 0;

Qy 3 HPOFEKXXXXXXXXXKSHQPK 25
Db 18 HPOFOAVNNVQGVVAHPTFK 40

RESULT 3
ID 08J219 PRELIMINARY; PRT; 663 AA.
AC 08J219;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)
DE Aminopectidase C.
GN Name=apcc;
OS Aspergillus niger.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;
OC Eucotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.
OX NCBI_TaxID=5061;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=N400;
RX MEDLINE=22459136; PubMed=12571053;
RX DOI=10.1128/AEM.69.2.1246-1250.2003;
RA Baeten D.E.J.W., Dekker P.J.T., Schaap P.J.;
RT "Aminopectidase C of Aspergillus niger is a novel phenylalanine
RT Aminopectidase.";
RL Appl. Environ. Microbiol. 69:1246-1250(2003).
DR EMBL; AJ316576; CAC87723.1; -.
DR MEROPS; S09.057; -.
DR GO; GO:0004177; F:aminopectidase activity; IEA.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR InterPro; IPR008262; Lipase AS.
DR InterPro; IPR001375; Peptidase_S9.
DR InterPro; IPR000379; Ser_estrs.
DR InterPro; IPR01042; To1B C.
DR InterPro; IPR01042; To1B C.
DR Pfam; PF00326; Peptidase_S9; 1.
DR PROSITE; PS00120; LIPASE_SER; UNKNOWN_1.
DR PROSITE; PS00300; SRP54; UNKNOWN_1.
KW Aminopectidase.
SQ SEQUENCE 663 AA; 72517 MW; C4EA59DA9743B545 CRC64;

Query Match          56.8%; Score 63; DB 2; Length 663;
Best Local Similarity 31.8%; Pred. No. 27;
Matches 7; Conservative 12; Mismatches 3; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXSHQ 22
Db 176 YSHPTFSPAGDRVSWIWNHPE 197

RESULT 4
Q6MKM3
```

```
ID 06MKM3 PRELIMINARY; PRT; 455 AA.
AC 06MKM3;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Hypothetical protein precursor.
GN OrderedLocustNames=Bq2365;
OS Bdellovibrio bacteriovorus.
OC Bacteria; Proteobacteria; Delaproteobacteria; Bdellovibrionales;
OC Bdellovibrionaceae; Bdellovibrio.
OX NCBI_TaxID=959;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HD100 / DSM 50701 / ATCC 15356 / NCIB 9529;
RX PubMed=14752164; DOI=10.1126/science.1093027;
RA Rendulic S., Jagtap P., Rosinus A., Eppinger M., Baar C., Lanz C.,
RA Keller H., Lambert C., Evans K.J., Goessmann A., Meyer F.,
RA Sockett R.E., Schuster S.C.;
RT "A predator unmasked: life cycle of Bdellovibrio bacteriovorus from a
RT genomic perspective.";
RL Science 303:689-692(2004).
DR EMBL; BX842652; CAE80184.1; -.
KW Complete proteome.
FT SIGNAL 1 11 Potential.
SQ SEQUENCE 455 AA; 51733 MW; 94F51E24304BDF9 CRC64;

Query Match          55.9%; Score 62; DB 2; Length 455;
Best Local Similarity 30.4%; Pred. No. 24;
Matches 7; Conservative 12; Mismatches 4; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXSHQ 23
Db 311 WTHPTFEDAVVSPFVETNSYF 333

RESULT 5
ID 09U6S0 PRELIMINARY; PRT; 421 AA.
AC 09U6S0;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2004 (TREMBLrel. 26, Last annotation update)
DE Calreticulin precursor.
GN Name=calret;
OS Strongylocentrotus purpuratus (Purple sea urchin).
OC Eukaryota; Metazoa; Echinodermata; Eleutherozoa; Echinozoa;
OC Echinoidea; Euechinoidea; Echinacea; Echinoidea; Strongylocentrotidae;
OC Strongylocentrotus.
OX NCBI_TaxID=7668;
RN [1]
RP SEQUENCE FROM N.A.
RA Susan J.M., Just M.L., Lennarz W.J.;
RT Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
CC -1-SIMILARITY: Belongs to the calreticulin family.
DR EMBL; AF177915; AAD55725.1; -.
DR HSSP; P18418; IK9C.
DR GO; GO:0005783; C:endoplasmic reticulum; IEA.
DR GO; GO:0005514; F:calcium ion storage activity; IEA.
DR GO; GO:0051082; F:unfolded protein binding; IEA.
DR GO; GO:006457; P:protein folding; IEA.
DR InterPro; IPR001580; Calret/calnex.
DR InterPro; IPR009169; Calreticulin.
DR InterPro; IPR009033; Calret calnex P.
DR InterPro; IPR008985; Cona_like_jec_g1.
DR InterPro; IPR000886; ER_target_S.
DR Pfam; PF00262; Calreticulin; 1.
DR PRINTS; PR00626; CALRETICULIN.
DR PROSITE; PS00804; CALRETICULIN_2; 1.
DR PROSITE; PS00805; CALRETICULIN_REPEAT; 3.
DR PROSITE; PS00014; ER_TARGET; UNKNOWN_1.
KW Chaperone; Endoplasmic reticulum; Signal.
FT SIGNAL 1 19 Potential.
```

```

FT CHAIN 20 421 calreticulin.
SQ SEQUENCE 421 AA; 48822 MW; 172C664F59F41P93 CRC64;

Query Match
Best Local Similarity 55.0%; Score 61; DB 2; Length 421;
Matches 7; Conservative 12; Mismatches 4; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXXXXSHPOF 23
Db 288 WYHPEIENSEYEADENLYSPSP 310

RESULT 6
O8CJD2 PRELIMINARY; PRT; 704 AA.
ID O8CJD2;
AC O8CJD2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DE 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Guanylyl cyclase alpha.1 subunit.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxId=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Nakamura I., Suzuki N.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB096020; BAC24016.1; -.
DR HSSP; P30803; IAS2.
DR GO; GO:0004383; F:guanylate cyclase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro; IPR001054; G_cyclase.
DR Pfam; PF00211; Guanylate_cyc; 1.
DR SMART; SM00044; CYCG; 1.
DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
KW lyase.
SQ SEQUENCE 704 AA; 79188 MW; F73FC97B68544E6A CRC64;

Query Match
Best Local Similarity 54.1%; Score 60; DB 2; Length 704;
Matches 8; Conservative 9; Mismatches 0; Indels 0; Gaps 0;

Qy 9 XXXXXXXXXXXXSHPOFEK 25
Db 688 GVDLVPRGSMWHPQFEK 704

RESULT 7
Q7TSS9 PRELIMINARY; PRT; 212 AA.
ID Q7TSS9;
AC Q7TSS9;
DT 01-OCT-2003 (TrEMBLrel. 25, Created)
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxId=10090;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=C57BL/6J; TISSUE=Mammary gland;
RC MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klanner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diachenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stalcen M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Uedlin T.B., Toehyuk S., Carninci P., Prange C.,

```

```

RA Raha S.S., Loguellano N.A., Peters G.J., Abramson R.D., Mullaly S.J.,
RA Bosak S.A., McEwen P.J., McKernan K.J., Malek J.A., Gunnarsson P.H.,
RA Richards S., Morley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA Villalón D.K., Wuzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,
RA Krzywinski M.I., Skalska U., Smailus D.E., Scherch A., Schein J.E.,
RA Jones S.J., Maira M.A.;
RT "Generation and initial analysis of more than 15,000 full-length human
RT and mouse cDNA sequences."
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903 (2002).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Mammary gland;
RA Strausberg R.;
RL Submitted (MAY-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; BC052931; AAH52931.1; -.
DR GO; GO:0005634; C:nucleus; IEA.
DR InterPro; IPR004665; Sp100.
DR Pfam; PF03172; Sp100; 1.
KW Hypothetical protein.
SQ SEQUENCE 212 AA; 24759 MW; 817B5D56E1A6FE36 CRC64;

Query Match
Best Local Similarity 52.3%; Score 58; DB 2; Length 212;
Matches 7; Conservative 12; Mismatches 4; Indels 0; Gaps 0;

Qy 3 HPOFEKXXXXXXXXXSHPOFEK 25
Db 130 HPSYEQGNTLCTHEKMAPCPFK 152

RESULT 8
Q27481 PRELIMINARY; PRT; 1113 AA.
ID Q27481;
AC Q27481;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Hypothetical protein C47E12.5.
GN Name=uba-1; ORFNames=C47E12.5;
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
OC Rhabditidae; Peloderinae; Caenorhabditis.
OX NCBI_TaxId=6239;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RC MEDLINE=99069613; PubMed=9851916;
RA none;
RT "Genome sequence of the nematode C.elegans: A platform for
RT investigating biology."
RL Science 282:2012-2018 (1998).
RN [2]
RP SEQUENCE FROM N.A.
RA STRAIN=Bristol N2;
RC STRAIN=Bristol N2;
RA Coles L.;
RL Submitted (JAN-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z68882; CA93101.1; -.
DR PIR; T20004; T20004.
DR HSSP; Q8TBC4; INGV.
DR WormBase; WBGene0006699; uba-1.
DR WormPeP; C47E12.5; CE05449.
DR GO; GO:0003824; F:catalytic activity; IEA.
DR GO; GO:0004839; F:ubiquitin activating enzyme activity; IEA.
DR GO; GO:0005512; P:ubiquitin cycle; IEA.
DR InterPro; IPR009036; Moeb.
DR InterPro; IPR000205; NAD_BS.
DR InterPro; IPR000594; Th1F domain.
DR InterPro; IPR000127; Ubact_repeat.
DR InterPro; IPR000011; Ugtin_activ_enz.

```

```
DR Pfam: PF00899; Trif; 2.  
DR Pfam: PF02334; UBACT; 2.  
DR PRINTS; PRO1849; UBIQUITINACT.  
DR TIGRFAMs; TIGR01408; Ubel; 1.  
DR PROSITE; PS00865; UBIQUITIN_ACTIVAT_2; 1.  
DR Hypothetical protein.  
SQ SEQUENCE 1113 AA; 124083 MW; 8D2198A97E4066CC CRC64;  
  
Query Match 52.3%; Score 58; DB 2; Length 1113;  
Best Local Similarity 34.8%; Pred. No. 2.3e+02;  
Matches 8; Conservative 11; Mismatches 4; Indels 0; Gaps 0;  
  
Qy 2 SHPOFEKXXXXXXXXXWHPQ 24  
Db 355 SHVPEKSLVPEPFIMDYAKE 377  
  
RESULT 9  
ID Q38575 PRELIMINARY; PRT; 73 AA.  
AC Q38575;  
DT 01-NOV-1996 (TREMBlrel. 01, Created)  
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)  
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
DE Srp.  
OS Bacteriophage HK022.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Siphoviridae;  
OC Lambda-like viruses.  
OX NCBI_TaxID=10742;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=92395669; PubMed=1522593;  
RA Atkinson B.L., Gottesman M.E.,  
RT "The E. coli ip080 mutation blocks antitermination by coliphage HK022  
RT Q-function.";  
RT J. Mol. Biol. 227:29-37(1992).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=20328598; PubMed=10860721; DOI=10.1006/jmbi.2000.3729;  
RA Junia R.U., Ford M.E., Duda R.L., Youlton A., Hatfull G.F.,  
RA Hendrix R.W.,  
RT "Genomic sequences of bacteriophages HK97 and HK022: pervasive genetic  
RT mosaicism in the lambdaoid bacteriophages.";  
RT J. Mol. Biol. 299:27-51(2000).  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Junia R., Ford M.E., Duda R.L., Youlton A., Hatfull G.F.,  
RA Hendrix R.W.,  
RT Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.  
DR EMBL; X60308; CAA42853.1; -.  
DR EMBL; AF069308; AAF30374.1; -.  
DR PIR; S28978; S28978.  
SQ SEQUENCE 73 AA; 8302 MW; AAC13F263A3F42CF CRC64;  
  
Query Match 51.4%; Score 57; DB 2; Length 73;  
Best Local Similarity 36.4%; Pred. No. 15;  
Matches 8; Conservative 11; Mismatches 3; Indels 0; Gaps 0;  
  
Qy 4 POFEXKXXXXXXXXXWHPQ 25  
Db 41 PDSDKNQLILIELTQYSHPPFEK 62  
  
RESULT 10  
ID Q7MMZ9 PRELIMINARY; PRT; 406 AA.  
AC Q7MMZ9;  
DT 01-MAR-2004 (TREMBlrel. 26, Created)  
DT 01-MAR-2004 (TREMBlrel. 26, Last sequence update)  
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
DE 2-poly(phenyl-6-methoxyphenol) hydroxylase.  
GN OrderedLocustNames=VVO918;  
OS Vibrio vulnificus (strain YJ016).
```

```
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrio.  
OX NCBI_TaxID=196600;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX PubMed=14656965; DOI=10.1101/gr.1295503;  
RA Chen C.-Y., Wu K.-M., Chang Y.-C., Chang C.-H., Tsai H.-C.,  
RA Liao T.-L., Liu Y.-M., Chen H.-J., Shen A.B.-T., Li J.-C., Su T.-L.,  
RA Shao C.-P., Lee C.-T., Hor L.-I., Tsai S.-F.,  
RT "Comparative genome analysis of Vibrio vulnificus, a marine  
RT pathogen.";  
RL Genome Res. 13:2577-2587(2003).  
DR EMBL; AP005333; BAC93682.1; -.  
DR GO; GO:0004497; P:monooxygenase activity; IEA.  
DR GO; GO:0006725; P:aromatic compound metabolism; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR000733; Flav_monooxygenase.  
DR InterPro; IPR002025; NAD_B5.  
DR InterPro; IPR003042; Rng_monooxygenase.  
DR InterPro; IPR010971; Ubl_OHases.  
DR Pfam; PF01360; Monooxygenase.1.  
DR PRINTS; PRO0420; RINGMONOXGNASE.  
DR TIGRFAMs; TIGR01988; Ubl-OHases; 1.  
DR Complete proteome.  
SQ SEQUENCE 406 AA; 45816 MW; 20A727877811CC98 CRC64;  
  
Query Match 51.4%; Score 57; DB 2; Length 406;  
Best Local Similarity 27.3%; Pred. No. 1e+02;  
Matches 6; Conservative 11; Mismatches 5; Indels 0; Gaps 0;  
  
Qy 1 WSHPOFEKXXXXXXXXXWHPQ 22  
Db 91 WQHQTATRCVCPYKRLTETHEPE 112  
  
RESULT 11  
ID Q8DFE9 PRELIMINARY; PRT; 406 AA.  
AC Q8DFE9;  
DT 01-MAR-2003 (TREMBlrel. 23, Created)  
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
DE 2-poly(phenyl-6-methoxyphenol) hydroxylase.  
GN OrderedLocustNames=VV10265;  
OS Vibrio vulnificus.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;  
OC Vibrionaceae; Vibrio.  
OX NCBI_TaxID=672;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=CMCP6;  
RA Rhee J.H., Kim S.Y., Chung S.S., Kim J.J., Moon Y.H., Jeong H.,  
RA Choy H.E.,  
RT "Complete genome sequence of Vibrio vulnificus CMCP6.";  
RT Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE016797; AAC08799.1; -.  
DR GO; GO:0004497; P:monooxygenase activity; IEA.  
DR GO; GO:0006725; P:aromatic compound metabolism; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR000733; Flav_monooxygenase.  
DR InterPro; IPR003042; Rng_monooxygenase.  
DR InterPro; IPR010971; Ubl_OHases.  
DR Pfam; PF01360; Monooxygenase.1.  
DR PRINTS; PRO0420; RINGMONOXGNASE.  
DR TIGRFAMs; TIGR01988; Ubl-OHases; 1.  
DR Complete proteome.  
SQ SEQUENCE 406 AA; 45880 MW; 3741DFC327ED37F CRC64;  
  
Query Match 51.4%; Score 57; DB 2; Length 406;  
Best Local Similarity 27.3%; Pred. No. 1e+02;  
Matches 6; Conservative 11; Mismatches 5; Indels 0; Gaps 0;  
  
Qy 1 WSHPOFEKXXXXXXXXXWHPQ 22
```

Db 91 WQHIOATRCVCPKRLTETHPE 112

## RESULT 12

064KA0 PRELIMINARY; PRT; 410 AA.  
AC 064KA0;  
DT 25-OCT-2004 (TREMBlrel. 28, Created)  
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)  
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)  
DE Calreticulin.  
OS Amblyomma brasiliense.  
OC Eukaryota; Metazoa; Arthropoda; Chelicerata; Arachnida; Acari;  
OC Parasitiformes; Ixodida; Ixodidae; Amblyomma.  
OX NCBI\_TaxID=257691;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Xu G., Fang Q.Q., Kelrns J.E., Durden L.A.;  
RL Submitted (SEP-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AY395247; AAR2933.1; -.  
SQ SEQUENCE 410 AA; 47241 MW; B6235DEFA8D378B1 CRC64;

Query Match 51.4%; Score 57; DB 2; Length 410;  
Best Local Similarity 22.7%; Pred. No. 1e+02;  
Matches 5; Conservative 14; Mismatches 3; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXWSHPQ 22

Db 288 WVHPEIDNPEYADPRLVHNP 309

## RESULT 13

090B81 PRELIMINARY; PRT; 207 AA.  
AC 090B81;  
DT 01-DEC-2001 (TREMBlrel. 19, Created)  
DT 01-DEC-2001 (TREMBlrel. 19, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE NeF.  
OS Name=nef;  
OS Human immunodeficiency virus 1.  
OC Viruses; Retroid viruses; Retroviridae; Lentivirus.  
OX NCBI\_TaxID=11676;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA MEDLINE=21637600; PubMed=11779357; DOI=10.1089/08992220175342103;  
RA Maestri T., Loubser S., Hide W., Hunt G., Morris L., Ramjee G.,  
RA Abdool-Karim S., Williamson C., Gray C.M.;  
RT "Conserved domains of subtype C nef from South African HIV type 1-  
RT infected individuals include cytotoxic T lymphocyte epitope-rich  
RT regions."  
RL Aids Res. Hum. Retroviruses 17:1681-1687(2001).  
RN (12)  
RP SEQUENCE FROM N.A.  
RA Maestri T.N., Loubser S., Hide W., Hunt G., Morris L., Ramjee G.,  
RA Abdool-Karim S., Williamson C., Gray C.;  
RL Submitted (JUL-2001) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AF397538; AAK98477.1; -.  
DR HSSP; P04324; 1ZEC.  
DR GO; GO:0005525; F:GTP binding; IEA.  
DR Pfam; PF00469; F-protein 1.  
DR ProDom; PD000031; HIV\_Nef; 1.  
KW Aids; Lipoprotein; Myristate.  
SQ SEQUENCE 207 AA; 23626 MW; FEC73E6A5D79D78 CRC64;

Query Match 50.5%; Score 56; DB 2; Length 207;  
Best Local Similarity 28.6%; Pred. No. 66;  
Matches 6; Conservative 12; Mismatches 3; Indels 0; Gaps 0;

Qy 5 QFEKXXXXXXWSHPQEK 25

Db 195 EFDSSLARHNLGPMKREFFK 205

## RESULT 14

09PJG8 PRELIMINARY; PRT; 759 AA.  
AC 09PJG8;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
DE General secretion pathway protein D.  
GN Name=gspD; Ordered locus names=TC0861;  
OS Chlamydia muridarum.  
OC Bacteria; Chlamydiae; Chlamydiales; Chlamydiaceae; Chlamydia.  
OX NCBI\_TaxID=83560;  
RN (1)  
RP SEQUENCE FROM N.A.  
RC STRAIN=Mopn / Nig9;  
RX MEDLINE=20150255; PubMed=10684935; DOI=10.1093/nar/28.6.1397;  
RA Read T.D., Brunham R.C., Shen C., Gill S.R., Heidelberg J.F.,  
RA White O., Hickey B.K., Peterson J.D., Uitterback T.R., Berry K.J.,  
RA Bass S., Linher K.D., Weidman J.F., Knout H.M., Craven B., Bowman C.,  
RA Dodson R.J., Gwinn M.L., Nelson W.C., DeBoy R.T., Kolonay J.F.,  
RA McClarty G., Salzberg S.L., Eisen J.A., Fraser C.M.;  
RL "Genome sequences of Chlamydia trachomatis Mopn and Chlamydia  
RT pneumoniae A839."  
RT Pneumoniae A839."  
RL Nucleic Acids Res. 28:1397-1406(2000).  
CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).  
CC -1- SIMILARITY: Belongs to the GSP D family.  
DR EMBL; AB02352; AAF39657.1; -.  
DR PIR; D81657; D81657.  
DR TIGR; TC0861; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0009306; P:protein secretion; IEA.  
DR InterPro; IPR001775; Bac\_GSPD.  
DR InterPro; IPR004846; GSPD/IIIProtein.  
DR InterPro; IPR001220; Nectin legB.  
DR InterPro; IPR005644; Nectin-like.  
DR Pfam; PF00263; Secretin\_1.  
DR Pfam; PF03958; Secretin\_N; 2.  
DR PRINTS; PR00811; BACTERIALGSPD.  
DR PROSITE; PS00307; LECTIN LEGUME\_BETA; UNKNOWN\_1.  
KW Complete proteome; Transport.  
SQ SEQUENCE 759 AA; 83771 MW; 7D5DDF55ED3FBA31 CRC64;

Query Match 50.5%; Score 56; DB 2; Length 759;  
Best Local Similarity 31.8%; Pred. No. 2.8e+02;  
Matches 7; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

Qy 1 WSHPOFEKXXXXXXWSHPQ 22

Db 103 WSAEVSXKGDPEYALVHNP 124

## RESULT 15

08LTT5 PRELIMINARY; PRT; 323 AA.  
AC 08LTT5;  
DT 01-OCT-2002 (TREMBlrel. 22, Created)  
DT 01-OCT-2002 (TREMBlrel. 22, Last sequence update)  
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
DE 5' to 3' exonuclease-like protein.  
OS Vibrio phage VPV262.  
OC Viruses; dsDNA viruses, no RNA stage; Caudovirales; Podoviridae;  
OC T7-like viruses.  
OX NCBI\_TaxID=194802;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA MEDLINE=2267917; PubMed=12781722;  
RA Hardies S.C., Comeau A.M., Serwer P., Suttle C.A.;  
RT "The complete sequence of marine bacteriophage VPV262 infecting vibrio  
RT parahaemolyticus indicates that an ancestral component of a T7 viral  
RT supergroup is widespread in the marine environment."  
RL Virology 310:359-371(2003).

DR EMBL; AY095314; AAM28378.1; -.  
DR GO; GO:0004567; F.DNA binding; IEA.  
DR GO; GO:0004567; F.exonuclease activity; IEA.  
DR InterPro; IPR008918; 5\_3\_exo\_C.  
DR InterPro; IPR000513; Exo\_N\_T.  
KW Exonuclease.  
KW SEQUENCE 323 AA; 37725 MW; 81CB297639B33

Query Match	50.0%;	Score 55.5;	DB 2;	Length 323;
Best Local Similarity	34.8%;	Pred. No. 1.3e+02;		
Matches	8;	Conservative	12;	Mismatches 2;
			Indels	1;
			Gaps	1;

```
QY      3 HPQFEKXXXXXXXXXXWSHPQFEK 25
          ||:|||:::~::~||~::~||
Db      78 HPEFEKRVADVVDIHFWIE-EFDK 99
```

Search completed: March 2, 2005, 12:44:28  
Job time : 94.8862 secs





CC controllable binding properties. (I) is particularly used for purifying  
 CC FP from dilute solution in batch formats (which use simpler apparatus  
 CC than column methods and result in lower loss of FP). The present sequence  
 CC is that of a streptavidin binding peptide disclosed with the invention  
 XX  
 SO Sequence 24 AA:

Query Match 89.6%; Score 99.5; DB 6; Length 24;  
 Best Local Similarity 96.0%; Pred. No. 6e-06;  
 Matches 24; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 1 WSHPOFEKXXXXXXXXXXWSHPQFEK 25  
 |||||  
 DB 1 WSHPOFEK-XXXXXXXXXXWSHPQFEK 24

RESULT 2  
 ABP60369

ID ABP60369 standard; peptide; 35 AA.

XX ABP60369;

XX 28-MAR-2003 (first entry)

XX Streptavidin binding peptide SEQ ID NO 10.

XX Streptavidin; protein chip; microtitre plate; detection.

XX Synthetic.

XX Key Location/Qualifiers

FT Misc-difference 9..27 /label= unknown

FT /note= "optionally deleted for 1-15 residues"

XX DE10113776-A1.

XX 02-OCT-2002.

XX 21-MAR-2001; 2001DE-01013776.

XX 21-MAR-2001; 2001DE-01013776.

XX (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

XX Schmidt T;

XX WPI; 2003-031166/03.

XX New isolated peptide, useful as affinity purification tag for recombinant  
 PT protein, comprises at least two high-affinity streptavidin-binding  
 PT modules.

XX Claim 7; Page 16; 18pp; German.

XX The invention relates to an isolated peptide (I) comprising at least two  
 CC individual modules separated by 0-50 amino acids, with each containing at  
 CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain  
 CC streptavidin binding modules, are useful as affinity handles for  
 CC purification of recombinant fusion proteins (FP), also for detecting FP,  
 CC e.g. on protein chips or microtitre plates. The modules in (I) bind  
 CC strongly to streptavidin, with a co-operative effect that provides  
 CC stronger binding than a single tag, but are displaced by a competitor.  
 CC (I) does not interfere with the function of attached proteins (II) (so it  
 CC may not be essential to remove it); facilitates detection and has easily  
 CC controllable binding properties. (I) is particularly used for purifying  
 CC FP from dilute solution in batch formats (which use simpler apparatus  
 CC than column methods and result in lower loss of FP). The present sequence  
 CC is that of a streptavidin binding peptide disclosed with the invention  
 XX

Sequence 35 AA;

Query Match 86.5%; Score 96; DB 6; Length 35;

Best Local Similarity 71.4%; Pred. No. 3e-05;  
 Matches 25; Conservative 0; Mismatches 0; Indels 10; Gaps 1;

QY 1 WSHPOFEK-----XXXXXXXXXXWSHPQFEK 25  
 |||||  
 DB 1 WSHPOFEKXXXXXXXXXXWSHPQFEK 35

RESULT 3

ID ABP60370 standard; peptide; 36 AA.

XX ABP60370;

XX 28-MAR-2003 (first entry)

XX Streptavidin binding peptide SEQ ID NO 11.

XX Streptavidin; protein chip; microtitre plate; detection.

XX Synthetic.

XX Key Location/Qualifiers

FT Region 9..28 /note= "GGGS repeats 2-5 optionally absent, residues 13-28"

FT 9..12 /label= GGGS\_repeat

XX DE10113776-A1.

XX 02-OCT-2002.

XX 21-MAR-2001; 2001DE-01013776.

XX 21-MAR-2001; 2001DE-01013776.

XX (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

XX Schmidt T;

XX WPI; 2003-031166/03.

XX New isolated peptide, useful as affinity purification tag for recombinant  
 PT protein, comprises at least two high-affinity streptavidin-binding  
 PT modules.

XX Claim 9; Page 16; 18pp; German.

XX The invention relates to an isolated peptide (I) comprising at least two  
 CC individual modules separated by 0-50 amino acids, with each containing at  
 CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain  
 CC streptavidin binding modules, are useful as affinity handles for  
 CC purification of recombinant fusion proteins (FP), also for detecting FP,  
 CC e.g. on protein chips or microtitre plates. The modules in (I) bind  
 CC strongly to streptavidin, with a co-operative effect that provides  
 CC stronger binding than a single tag, but are displaced by a competitor.  
 CC (I) does not interfere with the function of attached proteins (II) (so it  
 CC may not be essential to remove it); facilitates detection and has easily  
 CC controllable binding properties. (I) is particularly used for purifying  
 CC FP from dilute solution in batch formats (which use simpler apparatus  
 CC than column methods and result in lower loss of FP). The present sequence  
 CC is that of a streptavidin binding peptide disclosed with the invention  
 XX

Sequence 36 AA;

Query Match 86.0%; Score 95.5; DB 6; Length 36;  
 Best Local Similarity 44.4%; Pred. No. 3.7e-05;  
 Matches 16; Conservative 9; Mismatches 0; Indels 11; Gaps 1;

QY 1 WSHPOFEK-----XXXXXXXXXXWSHPQFEK 25  
 |||||  
 DB 1 WSHPOFEKGGSGGGSGGGSGGGSGWSHPQFEK 36

ID	ADAMS20243	standard; protein; 845 AA.
XX	ADAMS20243;	
XX	18-NOV-2004	(first entry)
XX	Human aggrecanase ADAMTS4 mutant E362Q protein with inserted Strep tag.	
XX	ADAMTS4, a disintegrin-like and metalloprotease;	
KW	thrombospondin type 1 motif 4; repolysin; zinc metalloprotease;	
KW	aggrecanase; osteopathic; antiinflammatory; antiarthritic; antirheumatic;	
KW	cytostatic; osteoarthritis; glioma; cancer; inflammatory joint;	
KW	rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;	
KW	human; enzyme; chromosome 1q21-q23; Strep tag; mutant; muteln.	
XX	Homo sapiens.	
OS	Synthetic.	
OS	Unidentified.	
XX	Key	Location/Qualifiers
FT	Misc-difference 362	/note="Wild-type Glu substituted by Gln"
FT	Misc-difference 520..527	/note="Strep tag peptide 2 inserted"
XX	WO2004011637-A2.	
XX	05-FEB-2004;	
PD	29-JUL-2003;	2003WO-US023484.
XX	29-JUL-2002;	2002US-0398721P.
XX	(AMHP) WYETH.	
PA	(CORC/) CORCORAN C J.	
PA	(PLAN/) FLANNERY C R.	
PA	(ZENG/) ZENG W.	
PA	(RACI/) RACIE L A.	
PA	(MCDO/) MCDONAGH T.	
PA	(PREE/) FREEMAN B A.	
PA	(GEOR/) GEORGIADIS K E.	
PA	(LAVA/) LAVALLIE E R.	
XX	Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T;	
PI	Freeman BA, Georgiadis KE, Lavallie ER;	
DR	WPI; 2004-143860/14.	
XX	New isolated, modified ADAMTS4 (aggrecanase) protein with improved	
PT	stability useful for identifying inhibitors of the enzyme activity for	
PT	treating aggrecanase-associated conditions, including osteoarthritis.	
XX	Claim 9; SEQ ID NO 40; 117pp; English.	
XX	The invention relates to a novel isolated, modified ADAMTS4 (a	
CC	thrombospondin-like and metalloprotease (repolysin type) with	
CC	thrombospondin type 1 motif 4) protein with improved stability compared	
CC	to a naturally occurring, full-length ADAMTS4 protein, where the modified	
CC	protein differs from the naturally-occurring, full-length ADAMTS4 protein	
CC	by at least one amino acid. ADAMTS proteins are a subfamily of zinc	
CC	metalloproteases and include aggrecanases amongst their members. The	
CC	protein of the invention demonstrates osteopathic, antiinflammatory,	
CC	antiarthritic, antirheumatic and cytostatic activities and may be useful	
CC	for treating aggrecanase-associated conditions, including osteoarthritis,	
CC	glioma, cancer inflammatory joint disease, rheumatoid arthritis, septic	
CC	arthritis, periodontal disease and Crohn's disease. The current sequence	
CC	is that of the human aggrecanase ADAMTS4 mutant E362Q protein with	
CC	inserted Strep tag of the invention.	

```

50 Sequence 845 AA;
Query Match 64.0%; Score 71; DB 8; Length 845;
Best Local Similarity 50.0%; Pred.No. 5.8;
Matches 9; Conservative 9; Mismatches 0; Indels 0; Gaps 0
OY 1 WSHPOFEKXXXXXXXXXX 18
|||||:|||||:|||||:
Db 520 WSHPOFEKXGGMGPWGPW 537

RESULT 5
AAW93967
ID AAW93967 standard; protein; 205 AA.
AC
AAW93967;
DT 02-JUL-1999 (first entry)
XX
DE Plasmd pBBP21 protein fragment.
XX
XX Lipocalin; anticalin; peptide loop; leaflet structure; antibody; therapy //;
KW tissue-specific surface molecule; tumour-specific surface molecule;
KM tumour imaging; bilin binding protein; Dbc protein.
XX
OS Synthetic.
XX
Key Location/Qualifiers
FH 1..21
FT /label= signal_peptide
FT Protein 22..205
FT /label= fusion_peptide
FT /note= "This fusion peptide is constructed from bilin
FT binding protein and a strep-Tag II region."
PN
XX WO916873-A1.
XX
XX 08-APR-1999.
XX
XX 25-SEP-1998; 98WO-DE002898.
XX
XX 26-SEP-1997; 97DE-01042706.
PR
PA (SKER/) SKERRA A.
PI
XX Skerra A, Beste G, Schmidt F, Stibora T;
XX WPI, 1999-255093/21.
DR N-PSDB; AAX24100.
XX
XX New anticalins produced by mutation of lipocalin family protein.
XX
XX Disclosure; Page 70-72; 80pp; German.
XX
XX This invention describes the preparation of anticalins from polypeptides
XX of the lipocalin family by mutation of amino acids in the region of the 4
XX peptide loops (at one end of the cylindrical 'leaflet' structure).
XX Anticalins can bind to a predetermined ligand with determinable affinity
XX anticalins or fusions of anticalins, are used: (a) when immobilized, for
XX separation; and (b) when labeled, for identification of ligands (or their
XX fusions or conjugates). Typically they bind to tissue- or tumor-specific
XX surface molecules and can be used for tumour imaging or directly for
XX tumour therapy. Mutations can be introduced into lipocalin polypeptides
XX more easily than into antibodies, since lipocalin polypeptides have only
XX 4 (contrast 6) sequence segments. The peptide loops can tolerate amino
XX acid changes without significant effect on folding. Anticalins have high
XX specific affinity, comparable with that for antibodies. This sequence
XX represents a fusion polypeptide used in the method of the invention
XX
XX Sequence 205 AA;
Query Match 61.3%; Score 68; DB 2; Length 205;
Best Local Similarity 39.1%; Pred.No. 2.2;

```



ID AAW93966 standard; protein; 396 AA.  
 AC AAW93966;  
 DT 02-JUL-1999 (first entry)  
 DE Plasmid pBBP20 protein fragment.  
 KW Lipocalin; anticalin; peptide loop; leaflet structure; antibody; therapy;  
 KW tissue-specific surface molecule; tumour-specific surface molecule;  
 XX tumour imaging; bilin binding protein; phage coat protein pIII.  
 OS Synthetic.  
 FH Key ;  
 FT Peptide 1..21  
 FT Peptide /label= signal\_peptide  
 FT Peptide 22..396  
 FT /label= fusion\_peptide  
 FT /note= "This fusion peptide is constructed from bilin  
 FT binding protein, a strep-Tag II region and a fragment of  
 FT phage coat protein pIII"  
 FT 22..195  
 FT /note= "Bilin binding protein"  
 FT Region 196..205  
 FT /note= "Strep-Tag II region"  
 FT Misc-difference 206  
 FT /note= "Encoded by TAG"  
 FT Protein 207..396  
 FT /note= "phage coat protein pIII"  
 FN WO9916873-A1.  
 PD 08-APR-1999.  
 PF 25-SEP-1998; 98WO-DE002898.  
 PR 26-SEP-1997; 97DE-01042706.  
 PA (SKERRA) SKERRA A.  
 PI Skerra A, Beese G, Schmidt F, Stidora T;  
 DR WP1; 1999-255093/21.  
 DR N-PSDB; AAX24099.  
 XX  
 XX New anticalin produced by mutation of lipocalin family protein.  
 PT  
 PS Disclosure; Page 67-69; 80pp; German.  
 CC This invention describes the preparation of anticalins from polypeptides  
 CC of the lipocalin family by mutation of amino acids in the region of the 4  
 CC peptide loops (at one end of the cylindrical 'leaflet' structure).  
 CC Anticalins can bind to a predetermined ligand with determinable affinity.  
 CC Anticalins or fusions of anticalins, are used: (a) when immobilized, for  
 CC separation; and (b) when labeled, for identification of ligands (or their  
 CC fusions or conjugates). Typically they bind to tissue- or tumor-specific  
 CC surface molecules and can be used for tumour imaging or directly for  
 CC tumour therapy. Mutations can be introduced into lipocalin polypeptides  
 CC more easily than into antibodies, since lipocalin polypeptides have only  
 CC 4 (contrast 6) sequence segments. The peptide loops can tolerate amino  
 CC acid changes without significant effect on folding. Anticalins have high  
 CC specific affinity, comparable with that for antibodies. This sequence  
 CC represents a fusion polypeptide used in the method of the invention  
 SO Sequence 396 AA;  
 Query Match 61.3%; Score 68; DB 2; Length 396;  
 Best Local Similarity 39.1%; Pred. No. 5.3;  
 Matches 9; Conservative 11; Mismatches 3; Indels 0; Gaps 0

```

Db      : | ::::::::::::::| |||||
          183 YSDFEAACKVNNNSMWSHPQPEK 205

RESULT 9
ID AAB46420 standard; protein; 396 AA.
AC AAB46420;
DT 06-APR-2001 (first entry)
DE Bilin binding-protein associated protein.
KW Bilin-binding protein; muten; BBP, digoxigenin.
OS Unidentified.
PN MO200075308-A1.
PD 14-DEC-2000.
PF 08-JUN-2000; 2000MO-DE001873.
PR 08-JUN-1999; 99DE-01026068.
PS (SKER/) SKERRA A.
PI Skerra A, Schlehuber S;
DR WPI: 2001-071071/08.
N-PSDB; AAF25695.

New muteins of bilin-binding protein, useful for detecting digoxigenin
being used as label in e.g. binding assays, are very selective for
digoxigenin.

Example 1; Page 55-56; 80pp; German.

This invention describes novel polypeptides (I) that are muteins of bilin
-binding protein (BBP), which can bind digoxigenin (Dig) or its
conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and
have an amino acid (aa) substitution at at least one of the positions 28,
31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its
fusion proteins, are used to bind, detect, determine, immobilize or
separate Dig or its conjugates with proteins, nucleic acids,
carbohydrates, other biological or synthetic macromolecules or low
molecular weight compounds, particularly in assays where Dig is being
used as a label. Compared with Dig-specific antibodies, (I) have a
specific structure and are easier to prepare. They have very high
specificity for Dig, relative to other steroids, and fusion partners may
be attached to either end without compromising their ability to bind
ligand

SQ Sequence 396 AA;
Query Match 61.3%; Score 68; DB 4; Length 396;
Best Local Similarity 39.1%; Pred. No. 5.3;
Matches 9; Conservative 11; Mismatches 3; Indels 0; Gaps 0;

OY 3 HPOFEKXXXXXXXXXXSMGHPQPEK 25
   : | ::::::::::::::| |||||
   : | ::::::::::::::| |||||
Db 183 YSDFEAACKVNNNSMWSHPQPEK 205

RESULT 10
ID AAB46423 standard; protein; 396 AA.
AC AAB46423;
DT 06-APR-2001 (first entry)

```



```

XX Example 4; Page 68-71; 80pp; German.
PS
CC This invention describes novel polypeptides (I) that are muteins of bilin
CC -binding protein (BBP), which can bind digoxigenin (Dig) or its
CC conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and
CC have an amino acid (aa) substitution at at least one of the positions 28,
CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its
CC fusion proteins, are used to bind, detect, determine, immobilize or
CC separate Dig or its conjugates with proteins, nucleic acids,
CC carbohydrates, other biological or synthetic macromolecules or low
CC molecular weight compounds, particularly in assays where Dig is being
CC used as a label. Compared with Dig-specific antibodies, (I) have a
CC simpler structure and are easier to prepare. They have very high
CC specificity for Dig, relative to other steroids, and fusion partners may
CC be attached to either end without compromising their ability to bind
CC ligand
XX
SQ Sequence 659 AA;
XX
Query Match 61.3%; Score 68; DB 4; Length 659;
Best Local Similarity 39.1%; Pred. No. 11;
Matches 9; Conservative 11; Mismatches 3; Indels 0; Gaps 0
OY 3 HPQFEKXXXXXXXXXXWSHPQFEK 25
: | :::::|||||
Db 637 YSDFSEACKVNNSNWSHPQFEK 659
RESULT 13
ID AAB46427 standard; protein; 659 AA.
XX
XX AAB46427;
XX
XX 06-APR-2001 (first entry)
DT
DE Bilin binding-protein associated protein #8.
XX
XX Bilin-binding protein; mutein; BBP; digoxigenin.
XX
XX Unidentified.
OS
XX WO200075308-A1.
XX
XX 14-DEC-2000.
PD
XX 08-JUN-2000; 2000WO-DE001873.
XX
XX 08-JUN-1999; 99DE-01026068.
XX
XX (SKER/) SKERRA A.
XX
XX Skerra A, Schlehuber S;
XX
XX WPI; 2001-071071/08.
XX
XX N-PSDB; AAF25712.
XX
XX New muteins of bilin-binding protein, useful for detecting digoxigenin
XX being used as label in e.g. binding assays, are very selective for
XX digoxigenin.
XX
XX Example 4; Page 72-74; 80pp; German.
XX
XX This invention describes novel polypeptides (I) that are muteins of bilin
XX -binding protein (BBP), which can bind digoxigenin (Dig) or its
XX conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and
XX have an amino acid (aa) substitution at at least one of the positions 28,
XX 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its
XX fusion proteins, are used to bind, detect, determine, immobilize or
XX separate Dig or its conjugates with proteins, nucleic acids,
XX carbohydrates, other biological or synthetic macromolecules or low
XX molecular weight compounds, particularly in assays where Dig is being
XX used as a label. Compared with Dig-specific antibodies, (I) have a
XX simpler structure and are easier to prepare. They have very high
XX specificity for Dig, relative to other steroids, and fusion partners may
XX be attached to either end without compromising their ability to bind
XX ligand
XX

```

[illegible]

xx	Sequence	248	AA;
SO	Query Match	60.4%	Score 67; DB 7; Length 248;
	Best Local Similarity	36.0%;	Pred. No. 3.9;
	Matches	9; Conservative	13; Mismatches 3; Indels 0; Gaps 0;
QY	1	WSHPQFEKXXXXXXXXXXWSHPQFEK	25
Db	192	WSHPQFEKPSLAELAKVLANRELDK	216

RESULT 15	
ADA27293	
ID	ADA27293 standard; protein; 258 AA.
XX	

DT 20-NOV-2003 (First entry)  
XX

DE Plasmid pNGAL7 fragment protein.  
XX

alpha2-microglobulin-related protein; A2m; mouse; 24p3; uterocalin.

OS Synthetic.

EH	Key	Location/Qualifiers
FT	Pentide	1 37

```

/note= "Ompra signal sequence"

```

Protein

Protein	Tag II and albumin binding domain"	Tag II and albumin binding domain"
FT	33	100
FT	33	100

FT FLOCEIN

```

199. :209
/note="Stron-Tac IT effect... res"

```

Protein

/note= "Albumin binding domain of Protein G"

PN WO2003029462-A1.

PD 10-APR-2003.

27-SEP-2001; 2001WO-EP011213.

27-SEP-2001; 2001WO-EP011213.

(PIER-) PIERIS PROTEOLAB AG.

Skerra A, Schlehuber S,

WPI; 2003-381639/36.  
N-Peptide; ADA37387

N-FCDB;

Generating a mutetin of a protein for validating the protein as drug target by subjecting the protein to mutagenesis at various positions.

corresponding to sequence positions of the human neutrophil gelatinase-associated lipocalin.

Disclosure; Page 59-61; 68pp; English.

The present invention relates to a method for generating mutants of human neutrophil gelatinase-associated lipocalin (hNGAL), rat alpha2-microglobulin-related protein (A $\alpha$ m) or mouse 24p3/urocalin, by subjecting the protein to mutagenesis. In the method of generating a mutant, a nucleic acid coding for the mutants of the protein resulting from mutagenesis is operably fused at the 3' end with a gene coding for the coat protein of a filamentous bacteriophage of the M13-family or for a fragment of the coat protein. The present sequence is the protein encoded by a fragment of plasmid pHNGAL7, used to illustrate the invention. This sequence comprises human hNGAL, a Strep-Tag II and an albumin binding domain of protein G.

```

xx      Sequence 258 AA;
sq
Query Match      60.4%; Score 67; DB 6; Length 258;
Best Local Similarity 36.0%; Pred. No. 4.1;
Matches 9; Conservative 13; Mismatches 3; Indels 0; Gaps 0
Qy      1 WSHPOFEKXXXXXXXXXXWSHPOFEK 25
      |||||
Db      202 WSHPOFEKPASIAEAKVIANNEIDK 226

```

Search completed: March 2, 2005, 13:02:49  
Job time : 113.602 secs

Search completed: March 2, 2005, 13:02:49  
Job time : 113.602 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 77.7439 Seconds  
(without alignments)  
105.489 Million cell updates/sec

Title: SEQ13  
Perfect score: 111  
Sequence: 1 wshpqfexkxxxxxxwshpqfex 25

Scoring table: BLOSUM62GX  
Gapop 10.0, Gapext 0.5

Searched: 1385339 seqs, 328044528 residues  
Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

```
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep:*
19: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
20: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99.5	89.6	24	US-10-026-578B-3	Sequence 3, Appl1
2	95.5	86.0	36	US-10-026-578B-10	Sequence 10, Appl1
3	82.5	74.3	36	US-10-026-578B-11	Sequence 11, Appl1
4	71	64.0	845	US-10-628-432-40	Sequence 40, Appl1
5	68	61.3	21	US-09-809-517A-30	Sequence 30, Appl1
6	68	61.3	22	US-09-809-517A-33	Sequence 33, Appl1
7	64.5	58.1	646	US-10-628-432-49	Sequence 49, Appl1
8	64.5	58.1	858	US-10-628-432-27	Sequence 27, Appl1
9	63	56.8	485	US-10-628-432-47	Sequence 47, Appl1
10	63	56.8	663	US-10-344-741-2	Sequence 2, Appl1
11	63	56.8	697	US-10-628-432-24	Sequence 24, Appl1
12	62.5	56.3	24	US-09-809-517A-31	Sequence 31, Appl1
13	62.5	56.3	25	US-09-809-517A-34	Sequence 34, Appl1

14	62.5	56.3	245	US-10-887-228A-1	Sequence 1, Appl1
15	62.5	56.3	246	US-10-887-228A-9	Sequence 9, Appl1
16	62.5	56.3	252	US-10-887-228A-5	Sequence 5, Appl1
17	62	55.9	117	US-09-977-137A-4	Sequence 4, Appl1
18	62	55.9	117	US-09-977-137A-5	Sequence 5, Appl1
19	62	55.9	117	US-09-977-137A-7	Sequence 7, Appl1
20	62	55.9	117	US-09-977-137A-8	Sequence 8, Appl1
21	62	55.9	117	US-09-977-137A-9	Sequence 9, Appl1
22	62	55.9	117	US-09-977-137A-10	Sequence 10, Appl1
23	62	55.9	117	US-09-977-137A-11	Sequence 11, Appl1
24	62	55.9	118	US-09-977-137A-12	Sequence 12, Appl1
25	62	55.9	118	US-09-977-137A-6	Sequence 6, Appl1
26	61	55.0	633	US-10-628-432-53	Sequence 53, Appl1
27	60.5	54.5	261	US-10-767-701-42119	Sequence 42119, A
28	60	54.1	24	US-10-026-578B-4	Sequence 4, Appl1
29	60	54.1	661	US-10-354-983-8	Sequence 8, Appl1
30	60	54.1	763	US-10-358-283-15	Sequence 15, Appl1
31	57.5	51.8	86	US-09-884-767A-9	Sequence 9, Appl1
32	56	50.5	87	US-10-424-599-272001	Sequence 272001, A
33	55	49.5	112	US-10-424-599-171374	Sequence 171374, A
34	55	49.5	126	US-10-437-963-142983	Sequence 142983, A
35	55	49.5	134	US-10-437-963-104065	Sequence 104065, A
36	55	49.5	158	US-10-424-599-184769	Sequence 184769, A
37	55	49.5	212	US-10-424-599-184075	Sequence 184075, A
38	55	49.5	252	US-10-437-963-200150	Sequence 200150, A
39	55	49.5	315	US-10-369-493-5628	Sequence 5628, A
40	55	49.5	318	US-10-219-810-46	Sequence 46, Appl1
41	55	49.5	318	US-10-380-132-22	Sequence 22, Appl1
42	55	49.5	318	US-10-471-243-20	Sequence 20, Appl1
43	55	49.5	330	US-10-219-810-47	Sequence 47, Appl1
44	54.5	49.1	143	US-10-437-963-196479	Sequence 196479, A
45	54.5	49.1	381	US-10-425-114-63501	Sequence 63501, A

#### ALIGNMENTS

```
US-10-026-578B-3
? Sequence 3, Application US/10026578B
? Publication No. US20030083474A1
? GENERAL INFORMATION:
? APPLICANT: IBA (GmbH)
? TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Tr
? FILE REFERENCE: 100810.01US1
? CURRENT APPLICATION NUMBER: US/10/026,578B
? PRIOR FILING DATE: 2002-11-11
? PRIOR APPLICATION NUMBER: DE 101 13 776.1
? PRIOR FILING DATE: 2001-03-21
? PRIOR APPLICATION NUMBER: PCT/EP01/11846
? PRIOR FILING DATE: 2001-10-12
? NUMBER OF SEQ ID NOS: 14
? SOFTWARE: Patentin version 3.1
? SEQ ID NO 3
? LENGTH: 24
? TYPE: PRT
? ORGANISM: Artificial sequence
? FEATURE:
? OTHER INFORMATION: Synthetic Peptide
? FEATURE:
? NAME/KEY: MISC FEATURE
? LOCATION: (3)..(9)
? OTHER INFORMATION: X represents a single amino acid at the position indicated
? FEATURE:
? NAME/KEY: MISC FEATURE
? LOCATION: (10)..(10)
? OTHER INFORMATION: X represents a single amino acid at the position indicated
? FEATURE:
? NAME/KEY: MISC FEATURE
? LOCATION: (11)..(11)
? OTHER INFORMATION: X represents a single amino acid at the position indicated
? FEATURE:
```

```
/ NAME/KEY: MISC FEATURE
/ LOCATION: (12)-(12)
/ OTHER INFORMATION: X represents a single amino acid at the position indicated
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (13)-(13)
/ OTHER INFORMATION: X represents a single amino acid at the position indicated
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (14)-(14)
/ OTHER INFORMATION: X represents a single amino acid at the position indicated
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (15)-(15)
/ OTHER INFORMATION: X represents a single amino acid at the position indicated
/ FEATURE:
/ NAME/KEY: MISC FEATURE
/ LOCATION: (16)-(16)
/ OTHER INFORMATION: X represents a single amino acid at the position indicated
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-3
```

```
Query Match      89.6%; Score 99.5; DB 14; Length 24;
Best Local Similarity 96.0%; Pred. No. 2.1e-05;
Matches 24; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
```

```
OY      1 WSHPOFEKXXXXXXXXXXWSHPQFEK 25
      |||||
Db      1 WSHPOFEK-XXXXXXXXXXWSHPQFEK 24
```

```
RESULT 2
US-10-026-578B-10
/ Sequence 10, Application US/10026578B
/ Publication No. US20030083474A1
/ GENERAL INFORMATION:
/ APPLICANT: IBA (GmbH)
/ APPLICANT: Schmidt, Thomas
/ TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
/ FILE REFERENCE: 100810.01US1
/ CURRENT APPLICATION NUMBER: US/10/026.578B
/ PRIOR FILING DATE: 2002-11-11
/ PRIOR APPLICATION NUMBER: DE 101 13 776.1
/ PRIOR FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: PCT/EP01/11846
/ NUMBER OF SEQ ID NOS: 14
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 10
/ LENGTH: 36
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Synthetic Peptide
/ NAME/KEY: MISC FEATURE
/ LOCATION: (9)-(28)
/ OTHER INFORMATION: X represents a single amino acid at each of the positions indicat
/ OTHER INFORMATION: ed, some of the amino acids may be missing. Where amino acids are
/ OTHER INFORMATION: missing, the total numbers of x will be no less than 5
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-10
```

```
Query Match      86.0%; Score 95.5; DB 14; Length 36;
Best Local Similarity 69.4%; Pred. No. 0.00011;
Matches 25; Conservative 0; Mismatches 0; Indels 11; Gaps 1;
```

```
OY      1 WSHPOFEK-----XXXXXXXXXXWSHPQFEK 25
      |||||
```

```
Db      1 WSHPOFEKXXXXXXXXXXXXXXXXXXWSHPQFEK 36
```

```
RESULT 3
US-10-026-578B-11
/ Sequence 11, Application US/10026578B
/ Publication No. US20030083474A1
/ GENERAL INFORMATION:
/ APPLICANT: IBA (GmbH)
/ APPLICANT: Schmidt, Thomas
/ TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
/ FILE REFERENCE: 100810.01US1
/ CURRENT APPLICATION NUMBER: US/10/026.578B
/ PRIOR FILING DATE: 2002-11-11
/ PRIOR APPLICATION NUMBER: DE 101 13 776.1
/ PRIOR FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: PCT/EP01/11846
/ PRIOR FILING DATE: 2001-10-12
/ NUMBER OF SEQ ID NOS: 14
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 11
/ LENGTH: 36
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Synthetic Peptide
/ NAME/KEY: MISC FEATURE
/ LOCATION: (9)-(28)
/ OTHER INFORMATION: Of the five (GGGS) repeats between position 9 and 28, one or more
/ OTHER INFORMATION: repeats may be missing. However, at least one (GGGS) repeat will
/ FEATURE:
/ NAME/KEY: misc feature
/ OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-11
```

```
Query Match      74.3%; Score 82.5; DB 14; Length 36;
Best Local Similarity 41.7%; Pred. No. 0.0053;
Matches 15; Conservative 9; Mismatches 1; Indels 11; Gaps 1;
```

```
OY      1 WSHPOFEK-----XXXXXXXXXXWSHPQFEK 25
      |||||
Db      1 WSHPOFEKGGSGGSGGSGGSGGSGGSGGSGGSHPOFEK 36
```

```
RESULT 4
US-10-628-432-40
/ Sequence 40, Application US/10628432
/ Publication No. US20040142863A1
/ GENERAL INFORMATION:
/ APPLICANT: Wyeth
/ TITLE OF INVENTION: Modified ADAMTS4 molecules
/ FILE REFERENCE: AM101378
/ CURRENT APPLICATION NUMBER: US/10/628.432
/ PRIOR FILING DATE: 2003-07-29
/ NUMBER OF SEQ ID NOS: 53
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 40
/ LENGTH: 845
/ TYPE: PRT
/ ORGANISM: Artificial
/ FEATURE:
/ OTHER INFORMATION: ADAMTS4 ASM with insertion
US-10-628-432-40
```

```
Query Match      64.0%; Score 71; DB 16; Length 845;
Best Local Similarity 50.0%; Pred. No. 7;
Matches 9; Conservative 9; Mismatches 0; Indels 0; Gaps 0;
```

```
OY      1 WSHPOFEKXXXXXXXXXXW 18
      |||||
Db      520 WSHPOFEKAGGMPGMPW 537
```

```

RESULT 5
US-09-809-517A-30
Sequence 30. Application US/09809517A
Patent No. US20020034733A1
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/proteins
TITLE OF INVENTION: particles via disulfide bonds
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
CURRENT FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.0
SEQ ID NO 30
LENGTH: 21
TYPE: PRP
ORGANISM: artificial sequence
FEATURES:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30

```

Query Match	61.3%	Score 68	DB 9	Length 21
Best Local Similarity	42.9%	Pred. No. 0.2		
Matches	9	Conservative 11	Mismatches 1	Indels 0
Qy	5	QFEKCKXXXXXXXXXSHPOFEK	25	
	:	: : : : : : : : : : : : : : :		
Db	1	EFDYKDDDDKGAFWSHPOFEK	21	

```

RESULT 6
US-09-809-517A-33
; Sequence 33, Application US/09809517A
; Patent No. US2002003473A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. US2002003473A1 method for displaying (poly)peptides/prote
; TITLE OF INVENTION: particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 22
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
; US-09-809-517A-33

```

```

Query Match 61.3%; Score 68; DB 9; Length 22;
Best Local Similarity 42.9%; Pzed No. 0.21;
Matches 9; Conservative 11; Mismatches 1; Indels 0;
Gaps 0;

Qy 5 QPEKXXXXXXXXXMSHPQEK 25
   : : : : : : : : : : : : : : : :
Db 2 EFDYDDDDDKGAPVMSHPQEK 22

```

RESULT 7  
US-10-628-432-49  
; Sequence 49, Application US/10628432

```

? Publication No. US20040142863A1
? GENERAL INFORMATION:
? APPLICANT: Myeth
? TITLE OF INVENTION: Modified ADAMTS4 molecules
? FILE REFERENCE: AM101378
? CURRENT APPLICATION NUMBER: US/10/628,432
? CURRENT FILING DATE: 2003-07-29
? NUMBER OF SEQ ID NOS: 53
? SOFTWARE: PatentIn version 3.1
? SEQ ID NO 49
? LENGTH: 646
? TYPE: PRT
? ORGANISM: Artificial
? FEATURE:
? OTHER INFORMATION: furin-processed construct E
US-10-628-432-49

Query Match          59.1%; Score 64.5; DB 16; Length 646;
Best Local Similarity 31.2%; Pred. NO. 35;
Matches 10; Conservative 12; Mismatches 3; Indels 7; Gaps 1;

```

```
Qy      1 WSHPQFE-----KXXXXXXXXXXWSHPQFEK 25
          | : : ::::::::::|
Db      615 WLHRRQAQLLEILRRPFWAGRGSAWSHPQFEK 646
```

```

RESULT 8
US-10-628-432-27
Sequence 27, Application US/10628432
Publication No. US20040142863A1
GENERAL INFORMATION:
APPLICANT: Wyeth
TITLE OF INVENTION: Modified ADAMTS4 molecules
FILE REFERENCE: AM01378
CURRENT APPLICATION NUMBER: US/10/628,432
CURRENT FILING DATE: 2003-07-29
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.1
SEQ ID NO 27
LENGTH: 858
TYPE: PRT
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: modified ADAMTS4 molecule
US-10-628-432-27

```

```

Query Match      58.1%  Score 64.5:  DB 16; Length 858;
Best Local Similarity 31.2%  Pred. No. 49;
Matches 10; Conservative 12; Mismatches 3; Indels 7; Gaps 1.

Oy 1 WSPQFE-----KXXXXXXXXXXWSPQFEK 25
    | | | | | : : : : : : : : : :
Db 827 WLRRAQILRLRLRRPWRGKGSAMWSPQFEK 858

```

```

1 RESULT 9
2 US-10-628-432-47
3 Sequence 47, Application US/10628432
4 Publication NO. US20040142863A1
5 GENERAL INFORMATION:
6 APPLICANT: Wyeth
7 TITLE OF INVENTION: Modified ADAMTS4 molecules
8 FILE REFERENCE: AM101378
9 CURRENT APPLICATION NUMBER: US/10/628,432
10 CURRENT FILING DATE: 2003-07-29
11 NUMBER OF SEQ ID NOS: 53
12 SOFTWARE: PatentIn version 3.1.1
13 SEQ ID NO 47
14 LENGTH: 465
15 TYPE: PRT
16 ORGANISM: Artificial
17 FEATURE:
18 OTHER INFORMATION: furin-processed construct C

```

US-10-628-432-47

Query Match 56.8%; Score 63; DB 16; Length 485;  
Best Local Similarity 36.7%; Pred. No. 38;  
Matches 11; Conservative 11; Mismatches 2; Indels 6; Gaps 1;

OY 2 SHPOFEK-----XXXXXXXXXXSHPOFEK 25  
DB 456 SKKFFDKMVCGGDGGSGGSAMSHPOFEK 485

RESULT 10  
US-10-344-741-2  
; Sequence 2, Application US/10344741  
; Publication No. US20040038371A1  
; GENERAL INFORMATION:  
; APPLICANT: Basten, Danielle  
; APPLICANT: Dekker, Petrus, Jacobus, Theodorus  
; APPLICANT: Schuurhuizen, Paul, William  
; APPLICANT: Schaap, Petrus, Johannes  
; APPLICANT: Visser, Jacob  
; APPLICANT: DSM NV  
; TITLE OF INVENTION: No. US20040038371A1 Amino-peptidase  
; FILE REFERENCE: 24615-20187.00  
; CURRENT APPLICATION NUMBER: US/10/344,741  
; PRIOR FILING DATE: 2001-08-22  
; PRIOR APPLICATION NUMBER: EP/NL 00202995.7  
; NUMBER OF SEQ ID NOS: 27  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 2  
; LENGTH: 663  
; TYPE: PRT  
; ORGANISM: Aspergillus niger N400

US-10-344-741-2

Query Match 56.8%; Score 63; DB 15; Length 663;  
Best Local Similarity 31.8%; Pred. No. 56;  
Matches 7; Conservative 12; Mismatches 3; Indels 0; Gaps 0;

OY 1 WSHPOFEKXXXXXXXXXXSHPO 22  
DB 176 YSHPTFSPAGDRVSWIQNHPE 197

RESULT 11  
US-10-628-432-24  
; Sequence 24, Application US/10628432  
; Publication No. US20040142863A1  
; GENERAL INFORMATION:  
; APPLICANT: Wyeth  
; TITLE OF INVENTION: Modified ADAMTS4 molecules  
; FILE REFERENCE: AM101378  
; CURRENT APPLICATION NUMBER: US/10/628,432  
; CURRENT FILING DATE: 2003-07-29  
; NUMBER OF SEQ ID NOS: 53  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 24  
; LENGTH: 697  
; TYPE: PRT  
; ORGANISM: Artificial  
; FEATURE:  
; OTHER INFORMATION: Truncated ADAMTS4 molecule  
US-10-628-432-24

Query Match 56.8%; Score 63; DB 16; Length 697;  
Best Local Similarity 36.7%; Pred. No. 59;  
Matches 11; Conservative 11; Mismatches 2; Indels 6; Gaps 1;

OY 2 SHPOFEK-----XXXXXXXXXXSHPOFEK 25  
DB 668 SKKFFDKMVCGGDGGSGGSAMSHPOFEK 697

RESULT 12  
US-09-809-517A-31  
; Sequence 31, Application US/09809517A  
; Patent No. US20020034733A1  
; GENERAL INFORMATION:  
; APPLICANT: Lohning, Corinna  
; TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote  
; FILE REFERENCE: MORPHO/11  
; CURRENT APPLICATION NUMBER: US/09/809,517A  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: EP 99114072.4  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: EP 00103551.8  
; PRIOR FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 31  
; LENGTH: 24  
; TYPE: PRT  
; ORGANISM: artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module  
US-09-809-517A-31

Query Match 56.3%; Score 62.5; DB 9; Length 24;  
Best Local Similarity 41.7%; Pred. No. 1.2;  
Matches 10; Conservative 11; Mismatches 0; Indels 3; Gaps 1;

OY 5 OFEK-----XXXXXXXXXXSHPOFEK 25  
DB 1 EFEOKLISEEDLNGAPWMSHPOFEK 24

RESULT 13  
US-09-809-517A-34  
; Sequence 34, Application US/09809517A  
; Patent No. US20020034733A1  
; GENERAL INFORMATION:  
; APPLICANT: Lohning, Corinna  
; TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote  
; FILE REFERENCE: MORPHO/11  
; CURRENT APPLICATION NUMBER: US/09/809,517A  
; PRIOR FILING DATE: 2001-03-15  
; PRIOR APPLICATION NUMBER: EP 99114072.4  
; PRIOR FILING DATE: 1999-07-20  
; PRIOR APPLICATION NUMBER: EP 00103551.8  
; PRIOR FILING DATE: 2000-02-18  
; NUMBER OF SEQ ID NOS: 41  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 34  
; LENGTH: 25  
; TYPE: PRT  
; ORGANISM: artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module  
US-09-809-517A-34

Query Match 56.3%; Score 62.5; DB 9; Length 25;  
Best Local Similarity 41.7%; Pred. No. 1.3;  
Matches 10; Conservative 11; Mismatches 0; Indels 3; Gaps 1;

OY 5 OFEK-----XXXXXXXXXXSHPOFEK 25  
DB 2 EFEOKLISEEDLNGAPWMSHPOFEK 25

RESULT 14  
US-10-887-228A-1  
; Sequence 1, Application US/10887228A  
; Publication No. US20050037402A1



*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 28.4553 Seconds  
(without alignments)  
65.585 Million cell updates/sec

Title: SEQ13

Perfect score: 111

Sequence: 1 wshpofekxxxxxxxxxwshpofek 25

Scoring table:

BLOSUM62DX  
Gapop 10.0, Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/prodata/1/1aa/5A\_COMB.pep:\*\n2: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*\n3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*\n4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*\n5: /cgn2\_6/prodata/1/1aa/6CTUS\_COMB.pep:\*\n6: /cgn2\_6/prodata/1/1aa/backfillseq1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	61.3	21	US-09-809-517A-30	Sequence 30, Appl
2	68	61.3	22	US-09-809-517A-33	Sequence 33, Appl
3	62.5	56.3	24	US-09-809-517A-31	Sequence 31, Appl
4	62.5	56.3	25	US-09-809-517A-34	Sequence 34, Appl
5	62	55.9	117	US-09-977-137A-5	Sequence 4, Appl
6	62	55.9	117	US-09-977-137A-5	Sequence 5, Appl
7	62	55.9	117	US-09-977-137A-7	Sequence 7, Appl
8	62	55.9	117	US-09-977-137A-8	Sequence 8, Appl
9	62	55.9	117	US-09-977-137A-9	Sequence 9, Appl
10	62	55.9	117	US-09-977-137A-10	Sequence 10, Appl
11	62	55.9	117	US-09-977-137A-11	Sequence 11, Appl
12	62	55.9	117	US-09-977-137A-12	Sequence 12, Appl
13	62	55.9	118	US-09-977-137A-6	Sequence 6, Appl
14	55	49.5	318	US-09-382-906A-2	Sequence 2, Appl
15	54	48.6	155	US-08-828-741B-11	Sequence 11, Appl
16	54	48.6	155	US-09-160-567-11	Sequence 11, Appl
17	54	48.6	155	US-09-710-299-11	Sequence 11, Appl
18	54	48.6	155	US-09-509-031-11	Sequence 11, Appl
19	54	48.6	495	US-08-828-741B-4	Sequence 4, Appl
20	54	48.6	495	US-09-160-567-4	Sequence 4, Appl
21	54	48.6	495	US-09-710-299-4	Sequence 4, Appl
22	54	48.6	495	US-09-809-517A-6	Sequence 6, Appl
23	53	47.7	10	US-09-809-517A-6	Sequence 6, Appl
24	53	47.7	83	US-09-732-210-1454	Sequence 1454, Ap
25	53	47.7	253	US-09-248-796A-21157	Sequence 21157, A
26	53	47.7	427	US-09-489-039A-11613	Sequence 11613, A
27	53	47.7	664	US-09-248-796A-16454	Sequence 16454, A

28	52.5	47.3	699	4	US-09-538-092-995	Sequence 995, App
29	52	46.8	146	4	US-09-270-767-35770	Sequence 35770, A
30	52	46.8	146	4	US-09-270-767-50987	Sequence 50987, A
31	52	46.8	243	4	US-09-252-991A-26513	Sequence 26513, A
32	52	46.8	646	4	US-09-252-991A-26479	Sequence 26479, A
33	51	45.9	8	3	US-08-948-097-2	Sequence 2, Appl
34	51	45.9	8	3	US-09-382-950-7	Sequence 7, Appl
35	51	45.9	8	3	US-09-382-736B-8	Sequence 8, Appl
36	51	45.9	8	4	US-09-619-103-9	Sequence 9, Appl
37	51	45.9	8	4	US-10-104-218-5	Sequence 5, Appl
38	51	45.9	8	4	US-09-809-517A-9	Sequence 9, Appl
39	51	45.9	155	4	US-08-887-534A-55	Sequence 55, Appl
40	51	45.9	155	4	US-09-527-431-55	Sequence 55, Appl
41	51	45.9	155	4	US-09-446-861-55	Sequence 55, Appl
42	51	45.9	604	3	US-09-586-935-3	Sequence 3, Appl
43	51	45.9	604	3	US-09-872-861-4	Sequence 4, Appl
44	51	45.9	754	4	US-09-198-452A-874	Sequence 874, App
45	51	45.9	754	4	US-09-438-185A-817	Sequence 817, App

#### ALIGNMENTS

```
RESULT 1
US-09-809-517A-30
Sequence 30, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
NUMBER OF SEQ ID NOS: 41
SOFTWARE: Patentin version 3.0
SEQ ID NO 30
LENGTH: 21
TYPE: PR
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30
Query Match
Best Local Similarity 61.3%; Score 68; DB 4; Length 21;
Matches 9; Conservative 11; Mismatches 1; Indels 0; Gaps 0;
QY 5 OPEKXXXXXXXXXWHPQFEK 25
:|:|||||
Db 1 EFDYKDDDKGAPWHPQFEK 21
RESULT 2
US-09-809-517A-33
Sequence 33, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
```

```

; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 22
; TYPE: PRT
; ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33

Query Match
Best Local Similarity 61.3%; Score 68; DB 4; Length 22;
Matches 9; Conservative 11; Mismatches 1; Indels 0; Gaps 0;

QY      5 QPEKXXXXXXXXXXWSHPQFEK 25
       :|:::|||||||
Db      2 EFDYDDDDKGAPWSPHQPFEK 22

RESULT 3
US-09-809-517A-31
Sequence 31, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
CURRENT FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.0
SEQ ID NO 34
LENGTH: 25
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-31

Query Match
Best Local Similarity 56.3%; Score 62.5; DB 4; Length 24;
Matches 10; Conservative 11; Mismatches 0; Indels 3; Gaps 1;

QY      5 QPEK--XXXXXXXXXXWSHPQFEK 25
       :|||:|||||||
Db      1 EFEOKLISEEDLNGAPWSPHQPFEK 24

RESULT 4
US-09-809-517A-34
Sequence 34, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on h
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
CURRENT FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.0
SEQ ID NO 34
LENGTH: 25
TYPE: PRT
ORGANISM: artificial sequence

```

```

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
- US-09-809-517A-34

Query Match          56.3%; Score 62.5; DB 4; Length 25;
Best Local Similarity 41.7%; Pred.No. 0.072;
Matches 10; Conservative 11; Mismatches 0; Indels 3; Gaps 1

QY      5 QPEK--XXXXXXXXXMSHPQFEK 25
       :||: ::::::::::::::|
Db      2 EFEOKLISEEDINGAWSHPQFEK 25

RESULT 5
US-09-977-137A-4
; Sequence 4, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4

Query Match          55.9%; Score 62; DB 4; Length 117;
Best Local Similarity 44.4%; Pred.No. 0.75;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

QY      8 KXXXXXXXMSHPQFEK 25
       ::::::::::::::|
Db     100 RKNVSCPSAWSHPQFEK 117

RESULT 6
US-09-977-137A-5
; Sequence 5, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-5

Query Match          55.9%; Score 62; DB 4; Length 117;
Best Local Similarity 44.4%; Pred.No. 0.75;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

```





```
/ TITLE OF INVENTION: Methods
/ FILE REFERENCE: 79-00
/ CURRENT APPLICATION NUMBER: US/09/977,137A
/ CURRENT FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,465
/ PRIOR FILING DATE: 2000-10-12
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 11
/ LENGTH: 117
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-11

Query Match          55.9%; Score 62; DB 4; Length 117;
Best Local Similarity 44.4%; Pred. No. 0.75;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

OY      8 KXXXXXXXXXSHPOFEK 25
Db      100 RKGWPCPSAWSHPOFEK 117

RESULT 12
US-09-977-137A-12
/ Sequence 12, Application US/09977137A
/ Patent No. 6750042
/ GENERAL INFORMATION:
/ APPLICANT: Summers, Anne O.
/ APPLICANT: Caguiat, Jonathan
/ TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
/ FILE REFERENCE: 79-00
/ CURRENT APPLICATION NUMBER: US/09/977,137A
/ PRIOR FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,465
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 12
/ LENGTH: 117
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-12

Query Match          55.9%; Score 62; DB 4; Length 117;
Best Local Similarity 44.4%; Pred. No. 0.75;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

OY      8 KXXXXXXXXXSHPOFEK 25
Db      100 RKGWPCPSAWSHPOFEK 117

RESULT 13
US-09-977-137A-6
/ Sequence 6, Application US/09977137A
/ Patent No. 6750042
/ GENERAL INFORMATION:
/ APPLICANT: Summers, Anne O.
/ APPLICANT: Caguiat, Jonathan
/ TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
/ FILE REFERENCE: 79-00
/ CURRENT APPLICATION NUMBER: US/09/977,137A
/ PRIOR FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,465
/ NUMBER OF SEQ ID NOS: 18
```

```
/ SOFTWARE: PatentIn Ver. 2.0
/ SEQ ID NO 6
/ LENGTH: 118
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-6

Query Match          55.9%; Score 62; DB 4; Length 118;
Best Local Similarity 44.4%; Pred. No. 0.76;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

OY      8 KXXXXXXXXXSHPOFEK 25
Db      101 RKGWPCPSAWSHPOFEK 118

RESULT 14
US-09-382-906A-2
/ Sequence 2, Application US/09382906A
/ Patent No. 6448475
/ GENERAL INFORMATION:
/ APPLICANT: Dellapenna, Dean
/ APPLICANT: Shintani, David
/ TITLE OF INVENTION: Manipulation of Tocopherol Levels in Transgenic Plants
/ FILE REFERENCE: 920905.90032
/ CURRENT APPLICATION NUMBER: US/09/382,906A
/ CURRENT FILING DATE: 1999-08-25
/ PRIOR APPLICATION NUMBER: 60/097,863
/ PRIOR FILING DATE: 1998-08-25
/ NUMBER OF SEQ ID NOS: 4
/ SOFTWARE: PatentIn Ver. 2.1
/ SEQ ID NO 2
/ LENGTH: 318
/ TYPE: PRT
/ ORGANISM: Synechocystis PCC6803
US-09-382-906A-2

Query Match          49.5%; Score 55; DB 4; Length 318;
Best Local Similarity 38.9%; Pred. No. 31;
Matches 7; Conservative 10; Mismatches 1; Indels 0; Gaps 0;

OY      6 FEKXXXXXXXXXSHPOF 23
Db      209 WEKPVKQLDQSHPAF 226

RESULT 15
US-08-828-741B-11
/ Sequence 11, Application US/08828741B
/ Patent No. 6043069
/ GENERAL INFORMATION:
/ APPLICANT: Koentgen, Frank
/ APPLICANT: Sues, Gabriele M.
/ APPLICANT: Tarlinton, David M.
/ APPLICANT: Treutlein, Herbert R.
/ TITLE OF INVENTION: CATALYTIC ANTIBODIES AND A METHOD OF
/ NUMBER OF SEQUENCES: 14
/ CORRESPONDENCE ADDRESSES:
/ ADDRESSEE: SCULLY, SCOTT, MURPHY & PRESSER
/ STREET: 400 Garden City Plaza
/ CITY: Garden City
/ STATE: New York
/ COUNTRY: United States of America
/ ZIP: 11530
/ COMPUTER READABLE FORM:
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: PatentIn Release #1.0, Version #1.30
/ CURRENT APPLICATION DATA:
```

APPLICATION NUMBER: US/08/828,741B  
FILING DATE: 26-MAR-1997  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: DIGILIO, Frank S.  
REGISTRATION NUMBER: 31,346  
REFERENCE/DOCKET NUMBER: 10591  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (516) 742-4343  
TELEFAX: (516) 742-4366  
TELEX: 230 901 SANS UR  
INFORMATION FOR SEQ ID NO: 11:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 155 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-828-741B-11

Query Match 48.6%; Score 54; DB 3; Length 155;  
Best Local Similarity 34.8%; Pred. No. 16;  
Matches 8; Conservative 10; Mismatches 5; Indels 0; Gaps 0;  
QY 1 WSHPOFERXXXXXXXXXWSHPOF 23  
DB 131 YSPGCTKLEIKRSGSAMRHPQF 153

Search completed: March 2, 2005, 12:25:37  
Job time: 28.4553 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using SW model

Run on: March 2, 2005, 12:20:44 ; Search time 19.8699 Seconds  
(without alignments)  
125.901 Million cell updates/sec

Title: SEQ14  
Perfect score: 112  
Sequence: 1 wshpqfexkxxxxxxxxxwshpqfex 26

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 183416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:\*  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	64	57.1	699	2	A54660	histidine rich cal
2	59	52.7	331	2	T51894	related to nucleot
3	59	52.7	771	2	A87532	glucosylase (impo
4	56	50.0	103	2	C85713	unknown protein en
5	56	50.0	103	2	A90904	hypothetical prote
6	56	50.0	1449	2	A12017	hypothetical prote
7	55.5	49.6	2037	2	S37178	fatty-acyl-CoA syn
8	55	49.1	98	2	E84598	probable auxin-reg
9	55	49.1	985	2	D82776	pyruvate dehydroge
10	55	49.1	1030	2	T16114	hypothetical prote
11	54	48.2	103	2	E90973	hypothetical prote
12	54	48.2	103	2	A85821	unknown protein en
13	54	48.2	363	2	T26842	hypothetical prote
14	54	48.2	389	2	T44957	heat shock protein
15	54	48.2	564	2	A70383	hydrogenase (EC 1.
16	54	48.2	585	2	S53072	glutamate decarbox
17	54	48.2	873	2	T25442	hypothetical prote
18	54	48.2	896	2	AB1156	conserved membrane
19	54	48.2	896	2	AE1514	conserved membrane
20	54	48.2	897	2	T43628	phosphatidylinosit
21	53.5	47.8	1028	2	T16475	hypothetical prote
22	53	47.3	645	2	I50680	alpha subunit of r
23	53	47.3	667	1	A48660	transketolase (EC
24	53	47.3	667	2	G91044	transketolase 2 is
25	53	47.3	667	2	B85889	transketolase 2 is
26	52	46.4	145	2	T19489	hypothetical prote
27	52	46.4	244	2	A39365	cyanamide hydratase
28	52	46.4	351	1	OKGMSA	protein kinase (EC
29	52	46.4	352	2	B82263	conserved hypothet

30	52	46.4	452	2	F84421	hypothetical prote
31	52	46.4	731	2	T29129	probable ATP/GTP b
32	52	46.4	882	2	G83018	pyruvate dehydroge
33	52	46.4	1218	2	S71376	glutamate receptor
34	51.5	46.0	142	2	B82223	hypothetical prote
35	51.5	46.0	358	2	T32238	hypothetical prote
36	51.5	46.0	443	2	AB2719	conserved hypothet
37	51.5	46.0	480	2	F87500	tolerance to colic
38	51.5	46.0	2157	2	A13009	peptide synthetase
39	51.5	46.0	2566	2	E98274	hypothetical prote
40	51	45.5	105	2	T10227	hypothetical prote
41	51	45.5	292	2	AB1724	transcription regu
42	51	45.5	292	2	A11353	transcription regu
43	51	45.5	301	1	S42101	protein kinase (EC
44	51	45.5	315	2	S18737	hypothetical prote
45	51	45.5	320	2	A00974	probable exported

## ALIGNMENTS

## RESULT 1

A54660 histidine rich calcium binding protein - human  
C:Species: Homo sapiens (man)  
C>Date: 02-Jun-1995 #sequence\_revision 02-Jun-1995 #text\_change 03-Jul-2004  
R: Hofmann, S.L.; Topham, M.; Hsieh, C.L.; Francke, U.  
Genomics 9, 656-669, 1991  
A:Title: cDNA and genomic cloning of HRC, a human sarcolemmal reticulum protein, and 1c  
A:Reference number: A54660; MUID:91244309; PMID:2037293  
A:Accession: A54660  
A:Status: preliminary  
A:Molecule type: mRNA  
A:Residues: 1-699 <HOF>  
A:Cross-references: UNIPROT:P23327; GB:M60052; NID:G183918; PIDN:AAA88071.1; PID:G183915  
C:Keywords: calcium binding

Query Match 57.1%; Score 64; DB 2; Length 699;  
Best Local Similarity 37.5%; Pred. No. 4.4;  
Matches 9; Conservative 10; Mismatches 5; Indels 0; Gaps 0;  
Oy 3 HPQFKXXXXXXXXXWSPQFK 26  
Db 67 HPDENKDVSTENGHHFMSHPDKK 90

## RESULT 2

T51894 related to nucleotide-binding protein [imported] - Neurospora crassa  
N:Alternate names: protein B23111.60  
C:Species: Neurospora crassa  
C>Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 03-Nov-2000  
R:Schulze, U.; Aign, V.; Hohlsel, J.; Brandt, P.; Fartmann, B.; Holland, R.; Nyakatura,  
C:Accession: T51894  
A:Title: cDNA and genomic cloning of HRC, a human sarcolemmal reticulum protein, and 1c  
A:Reference number: A54660; MUID:91244309; PMID:2037293  
A:Accession: T51894  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-331 <SCH>  
A:Cross-references: EMBL:AL391572; GSPDB:GN00116; NCSP:B23111.60  
A:Experimental source: BAC clone B23111; strain OR74A  
C:Genetics:  
A:Gene: NCSP:B23111.60  
A:Map position: 6  
A:introns: 34/3; 60/3; 87/1  
C:Superfamily: conserved probable membrane protein Y1L003W

```
Query Match Similarity      52.7%; Score 59; DB 2; Length 331;
                        Pred: 28.0%; No. 9.1;
Matched Local              7; Conservative 11; Mismatches 7; Indels 0; Gaps 0;
```

```
RESULT 3
A87532      glucoamylase [imported] - Caulobacter crescentus
C/Species:  Caulobacter crescentus
C/Date:      20-Apr-2001 #sequence_revision 20-Apr-2001 #text_change 09-Jul-2004
C/Accession: A87532
```

A:Name: *Cladobotryella*  
A:Accession: AF072499, MIM:608243, PMID:11259647  
A:Title: Complete Genome Sequence of *Cladobotryella*  
A:Reference number: AF072499, MIM:608243, PMID:11259647  
A:Accession: AF072499, MIM:608243, PMID:11259647  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-771 <STO>  
A:Cross-references: NID:gl3423797, PIDN:AAK4253.1, GSPDB:G  
C:Genetics:  
A:Gene: CCG282  
C:Superfamily: Cladobotryella glucan 1,4-alpha-glucosidase

```

Query Match      52.7%; Score 59; DB 2; Length 771,
Best Local Similarity 28.6%; Pred.No. 24,
Matches 6; Conservative 12; Mismatches 3; Indels 0; Gaps 0;

QY      1 WSHPQFEKXXXXXXXXXXMSH 21
      ||| :: : :: :: :: :: ::
Db      741 WSHAEYVKLRVSDGQVMDH 761

```

Unknown protein encoded within prophage CP-9330 [Imported] - *Escherichia coli* (strain O157:H7)  
C/Species: *Escherichia coli*  
C/Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C/Accession: C85713  
P/Name, N.T.: Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grobbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouotis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A/Title: Genome sequence of enterohaemorrhagic *Escherichia coli* O157:H7.  
A/Reference number: A85460; MUID:21074935; PMID:11206551  
A/Accession: C85713

A;Cross-references: UNIPROT:Q8X549; GB:AE005174; NID:G12315046; PIDN:AAG56167.1; GSPDB:G  
A;Experimental source: strain O15:H7, substrain EDL933  
C;Genetics:  
A;Gene: Z2097

```

Query Match      50.0%; Score 56; DB 2; Length 103;
Best Local Similarity 22.7%; Pred. No 6,1;
Matches 5; Conservative 16; Mismatches 1; Indels 0; Gaps 0;

5 QFEKXXXXXXXXXXXXXMSHPQEK 26
.: : : : : : : : : : : : : : : :
db 3 EIKEMPVVRDGGYGVWTHPEYER 24

```

```

RESULT 5
A90904
hypothetical protein Ecs2201 [imported] - Escherichia coli (strain O157:H7, substrain R)
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004

```

C/Accession: A90904  
R/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
Gasanara, N.; Yasunaga, T.; Kuwara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A/Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and genome  
A/Reference number: A96629; MUID:21156231; PMID:11258796  
A/Accession: A90904  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-103 <HAV>  
A/Cross-references: UNIROT:Q8X549; GB:BA000007; PIDN:BA835624.1; PID:913361667; GSPDB:G.  
A/Experimental source: strain O157:H7, substrain RIMD 0509952  
C/Genetics:  
A/Gene: ECs2201

```

QY      5 QFXXXXXXXXXXMSHPOPEX 26
      : : : : : : : : : : : : : :
Db      3 EIKEMPVNRDGYGWTHTPEYEK 24

Query Match      50.0%; Score 56; DB 2; length 103;
Best Local Similarity 22.7%; Pred. No. 6.1;
Matches      5; Conservative 16; Mismatches 1; Indels 0; Gaps 0;

```

RESULT 6  
AI2017  
hypothetical protein all1695 [imported] - Nostoc sp. (strain PCC 7120)  
C:Species: Nostoc sp. strain PCC 7120  
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120  
C:Date: 14-Dec-2001 #sequence\_revision 14-Dec-2001 #ext\_change 09-Jul-2004  
Accession: AI2017

R:Kanehisa, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasano, S.; Watanabe, A.; Iriyuchi, N.; Shimizu, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Tabata, S.  
DNA Res. 6, 205-213, 2001  
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena dactyloides* AT2017  
A:Reference number: AB1807, NCBI:21595285, PMID:1175840  
A:Status: preliminary  
A:Molecule: rRNA, rDNA

Query Match 50.0%; Score 56; DB 2; Length 1449;  
Best local similarity 38.0%;  
A:Residues: 1-1449 <KUR>  
A:Cross-references: UNIPROT:Q8W0C0; GB:BA000019; PIDR:BBB78061.1; PID:gi17135515\_GSPDB:G  
A:Experimental source: strain PCC 7120  
C:Genetics:  
A:Gene: all1695

Matches	6; Conservative	12; Mismatches	3; Indels	0; Gaps	0;
Qy	1	MSHPOPEKXXXXXXXXXXWSH	21		
Db	809	WHQQQGFQDPEVYTPLEKWSH	829		

S37178  
 fatty-acyl-CoA synthase (BC 2.3.1.86) beta chain - yeast (candida albicans)  
 N:Alternate names: fatty-acid synthase beta chain  
 C:Species: Candida albicans  
 C>Date: 10-Dec-1993 #sequence\_revision 10-Nov-1995 #text\_change 09-Jul-2004  
 C:Accession: S37178  
 R:Zhao, X.J.; Chlart, R.L.

A:Description: Isolation and sequence analysis of the *Candida albicans* FAS1 gene.  
A:Reference number: S37178  
A:Accession: S37178  
A:Molecule type: DNA  
A:Residues: 1-2037 <ZHA>  
A:Cross-References: UNIPROT:P34731; EMBL:X74952; NID:G402176; PIDN:CMA52907.1; PID:G402176metcsl

A:Residues: 1-2037 <ZHA>  
A:Cross-references: UNIPROT:P34731; EMBL:X74952; NID:g402176; PIDD:CAA52907.1; PID:g4021  
C:Genetics:  
A:Gene: FAS1  
C:Superfamily: yeast fatty-acyl-CoA synthase beta chain  
K:Keywords: acyltransferase; coenzyme A

```

Query Match          49.6%; Score 55.5; DB 2; Length 2037;
Best Local Similarity 31.8%; Pred. No. 2.4e+02;
Matches 7; Conservative 12; Mismatches 2; Indels 1; Gaps 1;

QY      3 HPOFEKXXXXXXXXXX-SHPQ 23
      |||:|||||:|||||:|||||:
Db      189 HPSFKITGTGILISMLKHP 210

RESULT 8
E84598
probable auxin-regulated protein [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Feb-2001 #sequence_revision 02-Feb-2001 #text_change 09-Jul-2004
C:Accession: E84598
R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.;
M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; Vanaken, S.E.; Umayam, L.; Tallon, L.;
euss, D.; Mierme, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.
Nature 402, 761-768, 1999
A>Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.
A:Reference number: A84420; MUID:20083487; PMID:10617197
A:Accession: E84598
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-98 <STO>
A:Cross-references: UNIPROT:Q9SKP2; GB:AE002093; NID:g4603922; PIDN:AAD29795.1; GSPDB:GN
C:Genetics:
A:Gene: At2g21210
A:Map position: 2
C:Superfamily: auxin-induced protein 10A

Query Match          49.1%; Score 55; DB 2; Length 98;
Best Local Similarity 33.3%; Pred. No. 8;
Matches 7; Conservative 12; Mismatches 2; Indels 0; Gaps 0;

QY      2 SHPOFEKXXXXXXXXXXMSHP 22
      |||:|||||:|||||:|||||:
Db      54 SHPOKILRKAEERGFDPH 74

RESULT 9
D82776
pyruvate dehydrogenase XPO669 [imported] - Xylella fastidiosa (strain 9a5c)
C:Species: Xylella fastidiosa
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 09-Jul-2004
C:Accession: D82776
R:Anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequen
Nature 406, 151-157, 2000
A>Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A:Reference number: A82515; MUID:20365717; PMID:10910347
A>Note: for a complete list of authors see reference number A59328 below
A:Accession: D82776
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-985 <SIM>
A:Cross-references: UNIPROT:Q9PF18; GB:AE003911; GB:AE003849; NID:g9105548; PIDN:AAF8347
A:Experimental source: strain 9a5c
R:Simpton, A.J.G.; Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; A
Briotes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carver, H
de-Neto, E.; Docena, C.; El-Porty, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A:Author: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm
J.D.; Junqueira, M.L.; Kemper, E.L.; Klatajma, J.P.; Krieger, J.E.; Kuramae, E.E.; Laizy
chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, E
A:Authors: Martins, E.M.F.; Matsukura, A.Y.; Menck, C.F.M.; Miranda, E.C.; Miyaki, C.Y.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmeri, D.A
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Sanceli, R.V.; Sawasak
A:Authors: da Silva, A.C.R.; da Silva, P.R.; da Silva, A.W.; Silva Jr., W.A.; da Silveir
M.; Tsunako, M.H.; Vallada, H.; Van Sluys, W.A.; Verjovski-Almeida, S.; Vettore, A.L.; Z
A:Reference number: A59328
A:Contents: annotation
C:Genetics:

```

```

A:Gene: XF0669
C:Superfamily: pyruvate dehydrogenase (lipoamide); thiamin pyrophosphate-binding domain

Query Match          49.1%; Score 55; DB 2; Length 985;
Best Local Similarity 29.6%; Pred. No. 1.2e+02;
Matches 8; Conservative 12; Mismatches 5; Indels 2; Gaps 1;

QY      1 WSHPOFEKXXXXXXXXXXS--HPOFE 25
      |||:|||||:|||||:|||||:
Db      854 WSCPSFSLRKDGFDTERMNLHPDSE 880

RESULT 10
T16114
hypothetical protein F20D12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 09-Jul-2004
C:Accession: T16114
R:Mu, X.
submitted to the EMBL Data Library, November 1995
A:Description: The sequence of C. elegans cosmid F20D12.
A:Reference number: Z18462
A:Accession: T16114
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-1030 <MX>
A:Cross-references: UNIPROT:Q19645; EMBL:U40933; NID:g1072142; PID:g1072148; PIDN:AAA816
C:Genetics:
A:Gene: CESP:F20D12.1
A:Introns: 159/1; 212/2; 263/3; 334/2; 385/1; 513/3; 754/2; 787/3; 906/3; 969/3; 999/2

Query Match          49.1%; Score 55; DB 2; Length 1030;
Best Local Similarity 33.3%; Pred. No. 1.2e+02;
Matches 7; Conservative 11; Mismatches 3; Indels 0; Gaps 0;

QY      4 POFEKXXXXXXXXXXMSHPQF 24
      |||:|||||:|||||:|||||:
Db      299 PDLKPSFVQGTLMNSDPHF 319

RESULT 11
E90973
hypothetical protein EC82757 [imported] - Escherichia coli (strain O157:H7, substrain R1
C:Species: Escherichia coli
C:Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: E90973
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E90973
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <HAY>
A:Cross-references: UNIPROT:Q8X4V0; GB:BA000007; PIDN:BA836180.1; PID:g13362225; GSPDB:C
A:Experimental source: strain O157:H7, substrain R1MD 0509952
C:Genetics:
A:Gene: EC82757

Query Match          48.2%; Score 54; DB 2; Length 103;
Best Local Similarity 25.0%; Pred. No. 12;
Matches 5; Conservative 15; Mismatches 0; Indels 0; Gaps 0;

QY      7 EKCKXXXXXXXXXXSHPOFEK 26
      |||:|||||:|||||:|||||:
Db      5 KEMPYERNBYGCMTHPEYEX 24

RESULT 12
A85821
unknown protein encoded within prophage CP-931U [imported] - Escherichia coli (strain O1
C:Species: Escherichia coli

```

C|Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C|Accession: AB58821  
R|Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Waller, L.; Grobbeck, E.J.; Davis, N.W.; Lam, A.; Dimalanta, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A|Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A|Reference number: AB5480; MUID:21074935; PMID:11206551  
A|Accession: AB58821  
A|Status: preliminary  
A|Molecule type: DNA  
A|Residues: 1-103 <STQ>  
A|Cross-references: UNIPROT:O8X4V0; GB:AE005174; NID:G12516136; PIDN:AMG57029.1; GSFDB:G  
A|Experimental source: strain O157:H7, substrain EDL933  
C|Genetics:  
A|Gene: Z3120

	Query Match	48.2%	Score 54	DB 2	Length 103;
	Best Local Similarity	25.0%	Pred No. 12;		
	Matches	5;	Conservative 15;	Mismatches 0;	Indels 0;
Qy	7	EKXXXXXXXXXXSHPOREX	26		
	:	: ::  :			
Db	5	KEMPERNEYGCMTHPEYEX	24		

```

RESULT 13
T26842
hypothetical protein Y43F4B.4 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 16-Aug-2004
C:Accession: T26842
R:Matthews, L.
submitted to the EMBL Data Library, January 1998
A:Reference number: Z20276
A:Accession: T26842
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-363 <MIL>
A:Cross-references: UNIPROT:Q45933; EMBL:AL021481; PIDs:CAA16333.1; CESP:Y43F4B.4
A:Experimental source: clone Y43F4B
C:Genetics:
A:Gene: CESP:Y43F4B.4
A:Introns: 22/3, 59/3, 80/2, 103/1, 232/2, 318/1
C:Superfamily: WD repeat homology

```

```

Query Match Similarity 48.2% Score 54; DB 2; Length 363,
Best Local Similarity 25.0%; Pred. No. 51,
Matches 6, Conservative 12, Mismatches 6, Indels 0, Gaps 0,
QY 1 WSHPOPEKXXXXXXXXXXSHPOF 24
      | : : : : : : : : : : : :
Db 52 WRSAAHWKCGAVRVIMAHPEF 75

```

```

RESULT 14
T44957
heat shock protein dnaJ [salinarity] - Halobacterium salinarum (ATCC 33170)
N/Alternate names: 40K chaperone, 40K heat shock protein
C/Species: Halobacterium salinarum
A/Variety: ATCC 33170
C/Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 09-Jul-2004
C/Accession: T44957
R/Bustard, K.; Gupta, R.S.
J. Mol. Evol. 45, 193-205, 1997
A>Title: The sequences of heat shock protein 40 (DnaJ) homologs provide evidence for a c
A/Reference number: Z22880, M0UD:97383250, PMID:9236279
A/Accession: T44957
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-389 <BUS>
A/Cross-references: UNIPROT:O34135, EMBL:U093357, PIDDN:AA896891.1
A/Experimental source: ATCC 33170
A/Note: the source is designated as Halobacterium cutribdum

```

```

C:Genetics:
A:Gene: dnaJ
C:Superfamily: heat shock protein dnaJ, dnaJ amino-terminal homology
F:4-68/Domain: dnaJ amino-terminal homology <DND>

Query Match      48.24; Score 54; DB 2; Length 389;
Best Local Similarity 33.34; Pred. No. 55;
Matches 7; Conservative 12; Mismatches 2; Indels 0;

QY      3  HPOFEKXXXXXXXXXXXXSHPO 23
      |||.....:||||
Db      276  HPDEFKDGDTISITATPSPFO 296

```

RESULT 15  
A70383  
hydrogenase (EC 1.18.99.1) 2 large chain - Aquifex aeolicus  
C|Species: Aquifex aeolicus  
C|Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jul-2004  
C|Accession: A70383  
R|Decker, G., Warren, P.V., Gaasterland, T., Young, W.G., Lenox, A.L., Graham, D.E., O'N.  
V.  
Nature 392, 353-358, 1998  
A|Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.  
A|Reference number: A70300; PMID:9819666; PMID:9537320  
A|Accession: A70383  
A|Status: Preliminary; nucleic acid sequence not shown; translation not shown  
A|Molecule type: DNA  
A|Residues: 1-564 <NOF>  
A|Cross-references: UNIPROT:O67092; GB:AB000715; NID:g2983460; PIDN:AAC07046.1; PID:g298  
A|Experimental source: strain VF5

A:Gene: mbnL2  
C:Function:  
A:Pathway: hydrogen metabolism  
A:Note: contains iron-sulfur and nickel  
C:Superfamily: hydrogenase (NiFe) large chain  
C:Keywords: hydrogen metabolism; iron; iron-sulfur protein; membrane bound; metalloprote  
F:60\_63\_546\_549/binding site: nickel (Cys) #status predicted  
F:63\_549/binding site: iron (Cys) #status predicted  
F:67/Active site: His #status predicted

	Query Match	48.2%	Score 54	DB 2	Length 564
	Best Local Similarity	23.8%	Pred. No. 85		
	Matches	5	Conservative 15	Mismatches 1	Indels 0
	Gaps	0			
QY	5 QFEKXXXXXXXXXXMSHPDFE	25			
	: : : : :   : : : : :				
DG	157 QVERGGGLGFNNKGYYMNHPEYK	177			

```
Search completed: March 2, 2005, 12:29:00
Job time : 20.8699 secs
```



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OW protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 96.6016 Seconds

(without alignments)  
137.824 Million cell updates/sec

Title: SEQ14  
Perfect score: 112  
Sequence: 1 whpogfekxxxxxxxxxwhpogfek 26

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt 03:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	64	57.1	699	1 SRCH HUMAN	P23327 homo sapien
2	62	55.4	722	2 Q7YSZ1	Q7YSZ1 maecca faec
3	61	54.5	704	2 Q8CDD2	Q8CDD2 rattus norv
4	61	54.5	743	2 Q8CH90	Q8CH90 rattus norv
5	60	53.6	782	2 Q9HGH1	Q9HGH1 pichia guil
6	59	52.7	253	2 Q7ZNB4	Q7ZNB4 leprospira
7	59	52.7	253	2 Q8FRC3	Q8FRC3 leprospira
8	59	52.7	309	2 Q7S6P7	Q7S6P7 neurospora
9	59	52.7	400	2 Q8FTK4	Q8FTK4 methanosarc
10	59	52.7	771	2 Q9A613	Q9A613 xanthomonas
11	58	51.8	895	2 Q8P4T1	Q8P4T1 xanthomonas
12	57.5	51.3	107	2 Q6Z315	Q6Z315 oryza sativ
13	57.5	51.3	380	2 Q742U3	Q742U3 mycobacteri
14	57	50.9	202	2 Q6E071	Q6E071 oryza sativ
15	57	50.9	1111	2 Q7SAB0	Q7SAB0 neurospora
16	56.5	50.4	270	2 Q9B1U4	Q9B1U4 leishmania
17	56	50.0	103	2 Q8X549	Q8X549 escherichia
18	56	50.0	192	2 Q7VBE9	Q7VBE9 prochloroco
19	56	50.0	299	2 Q9E7M7	Q9E7M7 rhizobium 1
20	56	50.0	408	2 Q7XNS9	Q7XNS9 oryza sativ
21	56	50.0	508	1 MATR_MARSC	Q8WKK9 marichnum s
22	56	50.0	508	2 Q9B8G0	Q9B8G0 vanoyenell
23	56	50.0	508	2 Q9B8G3	Q9B8G3 oserya coul
24	56	50.0	577	2 Q86X18	Q86X18 homo sapien
25	56	50.0	640	1 HTPG_NITRU	Q8ZLV8 nitrosomona
26	56	50.0	1449	2 Q8YWC0	Q8YWC0 anabaena sp
27	55.5	49.6	409	2 Q7QHY8	Q7QHY8 anopheles g
28	55.5	49.6	2037	1 FAST_CANAL	P34731 c fatty aci
29	55	49.1	98	2 Q9SKP2	Q9SKP2 arabidopsis
30	55	49.1	181	2 Q8MXH8	Q8MXH8 plodia inte
31	55	49.1	295	2 Q949H3	Q949H3 hevea brasl

32	55	49.1	295	2 Q8GUD7	Q8GUD7 hevea brasl
33	55	49.1	354	2 Q8ZNS5	Q8ZNS5 pyrobaculum
34	55	49.1	381	2 Q7Q9L2	Q7Q9L2 anopheles g
35	55	49.1	847	2 Q8EEC3	Q8EEC3 shewanella
36	55	49.1	985	2 Q9PFI8	Q9PFI8 xyella fas
37	55	49.1	1030	2 Q19645	Q19645 caenorhabdi
38	55	49.1	4903	1 MLJ3_MOUSE	Q8BZH4 mus musculu
39	55	49.1	4911	1 MLJ3_HUMAN	Q8BZH4 mus sapien
40	54.5	48.7	334	2 Q6ZBS0	Q6ZBS0 burkholderi
41	54.5	48.7	334	2 Q6ZBL9	Q6ZBL9 burkholderi
42	54.5	48.7	471	2 Q8NTJ1	Q8NTJ1 corynebacte
43	54.5	48.7	603	2 Q8J219	Q8J219 aspergillus
44	54	48.2	103	2 Q8X4V0	Q8X4V0 escherichia
45	54	48.2	208	2 Q6ZTR3	Q6ZTR3 bacillus 11

## ALIGNMENTS

```

RESULT 1
ID SRCH_HUMAN STANDARD; PRT; 699 AA.
AC P23327;
DT 01-NOV-1991 (Rel. 20, Created)
DT 01-NOV-1991 (Rel. 20, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Sarcoplasmic reticulum histidine-rich calcium-binding protein
DE Precursor.
GN Name=HRC; Synonyms=HCP;
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A., AND VARIANTS.
RC TISSUE=Skeletal muscle;
RX MEDLINE=91244309; PubMed=2037293;
RA Hofmann S.L., Topham M., Hsieh C.-L., Franke U.;
RT "cDNA and genomic cloning of HRC, a human sarcoplasmic reticulum
RT protein, and localization of the gene to human chromosome 19 and mouse
RT chromosome 7."
RL Genomics 9:656-669(1991).
CC -!- FUNCTION: May play a role in the regulation of calcium
CC sequestration or release in the SR of skeletal and cardiac muscle.
CC -!- SUBCELLULAR LOCATION: Sarcoplasmic reticulum lumen.
CC -!- SIMILARITY: Strong, to rabbit HRC.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL; M60052; AAA88071.1; -.
DR PIR; A54660; A54660.
DR GeneW; HGNC:5178; HRC.
DR MIM; 142705; -.
DR GO; GO:0005509; F:calcium ion binding; TAS.
DR GO; GO:0006936; P:muscle contraction; TAS.
KW Calcium-binding; Polymorphism; Repeat; Signal.
FT SIGNAL 1 28
FT CHAIN 29 699
FT FT
FT DOMAIN 193 204 Sarcoplasmic reticulum histidine-rich
FT DOMAIN 246 261 calcium-binding protein.
FT DOMAIN 106 365 Glu-rich (acidic).
FT REPEAT 180 213 4 x tandem repeats, acidic.
FT REPEAT 238 270 1-1.
FT REPEAT 295 318 1-2.
FT REPEAT 343 365 1-3.
FT DOMAIN 106 342 6 X approximate tandem repeats.

```

```

FT REPEAT 106 121 2-1.
FT REPEAT 134 154 2-2.
FT REPEAT 135 177 2-3.
FT REPEAT 214 237 2-4.
FT REPEAT 271 294 2-5.
FT REPEAT 319 342 2-6.
FT DOMAIN 627 673 Metal-binding (potential).
FT VARIANT 96 96 S -> A (in dbSNP:3745297).
FT /FTID=VAR_005623.
FT VARIANT 204 204 Missing.
FT /FTID=VAR_011622.
SQ SEQUENCE 699 AA; 80244 MW; 9922EBDF012C61D CRC64;

Query Match
Best Local Similarity 57.1%; Score 64; DB 1; Length 699;
Matches 9; Conservative 10; Mismatches 5; Indels 0; Gaps 0;

QY 3 HPOFEKXXXXXXXXXSHPOFEK 26
DB 67 HPDENKDVSTENGHFWSHPDRK 90

RESULT 2
QYYS21 PRELIMINARY; PRT; 722 AA.
AC QYYS21;
DT 01-OCT-2003 (TREMBlrel. 25, Created)
DT 01-OCT-2003 (TREMBlrel. 25, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Histidine-rich calcium binding protein.
GN Name=HRC;
OS Macaca fascicularis (Grab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;
OC Cercopithecoidea; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Heart;
RA Hong S., Cho C.;
RL Submitted (May-2003) to the EMBL/GenBank/DBJ databases.
DR EMBL; A305856; AAF70486.1; -.
DR InterPro; IPR001450; 4Fe4S_ferredoxin.
DR PROSITE; PS00198; 4Fe4S_FERREDOXIN; UNKNOWN_1.
DR PROSITE; PS00328; HCP; UNKNOWN_1.
SQ SEQUENCE 722 AA; 82753 MW; 87CBE737FA647F24 CRC64;

Query Match
Best Local Similarity 55.4%; Score 62; DB 2; Length 722;
Matches 8; Conservative 12; Mismatches 5; Indels 0; Gaps 0;

QY 2 SHPOFEKXXXXXXXXXSHPOFEK 26
DB 66 NHPDGNKDVSTENGHFWSHPDRK 90

RESULT 3
Q8CJD2 PRELIMINARY; PRT; 704 AA.
AC Q8CJD2;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Guanylyl cyclase alpha 1 subunit.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Nakamura T., Suzuki N.;
RL Submitted (Nov-2002) to the EMBL/GenBank/DBJ databases.

```

```

DR EMBL; AB096020; BAC24016.1; -.
DR HSSP; P30803; IAZS.
DR GO; GO:0004383; F:guanylate cyclase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro; IPR01054; G_cyclase.
DR Pfam; PF00211; Guanylate_cyc; 1.
DR SMART; SM00044; CYC; 1.
DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
SQ SEQUENCE 704 AA; 79188 MW; F73FC97B68544E6A CRC64;

Query Match
Best Local Similarity 54.5%; Score 61; DB 2; Length 704;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

QY 9 XXXXXXXXXXXXSHPOFEK 26
DB 687 SGVDLVRGSMWSHPQFEK 704

RESULT 4
Q8CH90 PRELIMINARY; PRT; 743 AA.
AC Q8CH90;
DT 01-MAR-2003 (TREMBlrel. 23, Created)
DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Soluble guanylyl cyclase alpha 2 subunit E219G mutant.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Kidney;
RA Nakamura T., Yao Y., Suzuki N.;
RL Submitted (Dec-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB097860; BAC44887.1; -.
DR HSSP; P30803; IAZS.
DR GO; GO:0004383; F:guanylate cyclase activity; IEA.
DR GO; GO:0016829; F:lyase activity; IEA.
DR GO; GO:0007242; P:intracellular signaling cascade; IEA.
DR InterPro; IPR01054; G_cyclase.
DR InterPro; IPR009080; ENAbyn_1a_bind.
DR Pfam; PF00211; Guanylate_cyc; 1.
DR SMART; SM00044; CYC; 1.
DR PROSITE; PS00452; GUANYLATE_CYCLASES_1; 1.
DR PROSITE; PS50125; GUANYLATE_CYCLASES_2; 1.
SQ SEQUENCE 743 AA; 83251 MW; ACF5C53E0982813A CRC64;

Query Match
Best Local Similarity 54.5%; Score 61; DB 2; Length 743;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

QY 9 XXXXXXXXXXXXSHPOFEK 26
DB 726 RETSLVRGSMWSHPQFEK 743

RESULT 5
Q9HGHI PRELIMINARY; PRT; 782 AA.
AC Q9HGHI;
DT 01-MAR-2001 (TREMBlrel. 16, Created)
DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)
DE Type II DNA topoisomerase (Fragment).
GN Name=top2;
OS Pichia guilliermondii (Yeast) (Candida guilliermondii).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;

```

CC Saccharomycetales; Saccharomycetaceae; Pichia.  
OX NCBI\_TaxID=4929;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NTM4;  
RX MEDLINE=21365449; PubMed=11470534; DOI=10.1016/S0978-1119(01)00526-1;  
RA Kato M., Ozeki M., Kikuchi A., Kanbe T.;  
RT "Phylogenetic relationship and mode of evolution of yeast DNA  
topoisomerase II gene in the pathogenic Candida species.";  
RL Gene 272:275-281(2001).  
CC -1- FUNCTION: Control of topological states of DNA by transient  
breakage and subsequent rejoining of DNA strands. Topoisomerase II  
makes double-strand breaks (by similarity).  
CC -1- CATALYTIC ACTIVITY: ATP-dependent breakage, passage and rejoining  
of double-stranded DNA.  
CC -1- SUBUNIT: Homodimer (By similarity).  
CC -1- MISCELLANEOUS: Eukaryotic topoisomerase I and II can relax both  
negative and positive supercoils, whereas prokaryotic enzymes  
relax only negative supercoils (by similarity).  
CC -1- SIMILARITY: Belongs to the type II topoisomerase family.  
DR HEMBL; AB049145; BABJ3754.1; -.  
DR HSSP; P06786; 1BYT.  
DR GO; GO:0005634; C:nucleus; IEA.  
DR GO; GO:0005554; F:ATP binding; IEA.  
DR GO; GO:0003918; F:DNA topoisomerase (ATP-hydrolyzing) activity; IEA.  
DR GO; GO:0006265; P:DNA topological change; IEA.  
DR InterPro; IPR003594; AtfBind\_ATPase.  
DR InterPro; IPR003957; CBFA\_NFYB\_topIs.  
DR InterPro; IPR011558; DNA\_Gyrase\_B.  
DR InterPro; IPR001241; DNA\_TopoisoltI.  
DR InterPro; IPR002205; DNA\_TopoisolvI.  
DR Pfam; PF00204; DNA\_gyraseB; 1.  
DR Pfam; PF0521; DNA\_topoisolvI; 1.  
DR Pfam; PF02518; HATPase\_C; 1.  
DR PRINTS; PR00615; CCAATSUBUNTA.  
DR PRODOM; PD149633; DNA\_gyrase\_B\_1.  
DR PRODOM; PD000742; DNA\_topoisolvI; 1.  
DR SMART; SMO0387; HATPase\_C; 1.  
DR SMART; SMO0433; TOP2c; 1.  
DR SMART; SMO0434; TOP2c; 1.  
KM ATP-binding; DNA-binding; Isomerase; Topoisomerase.  
FT NON\_TER 1  
FT NON\_TER 1  
FT FT 782 782  
SQ SEQUENCE 782 AA; 88612 MW; 0B9CFDB3BDABA6D CRC64;

Query Match 53.6%; Score 60; DB 2; Length 782;  
Best Local Similarity 28.0%; Pred.No. 1.4e+02;  
Matches 7; Conservative 14; Mismatches 4; Indels 0; Gaps 0.

OY 1 WSHPOFEKXXXXXXKXMSHPQF 25  
::|::|::|::|::|::|::|::|:  
Db 544 YNMPDEFKRETEGTCMKIKHYFK 568

RESULT 6  
O72NB4 PRELIMINARY; PRT; 253 AA.  
AC O72NB4.  
DT 05-JUL-2004 (TRMBLRel. 27, Created)  
DT 05-JUL-2004 (TRMBLRel. 27, Last sequence update)  
DE 05-JUL-2004 (TRMBLRel. 27, Last annotation update)  
DI dipeptide/cisloleucine ABC transport system permease protein.  
GN OrderedLocustNmeas-LIC12953;  
OS Leptospira interrogans (Betagroup Icterohaemorrhagiae / serovar Copenhagen1).  
OC Bacteria; Spirochaetes; Spirochaetales; Leptospiiraceae; Leptospira.  
OX NCBI\_TaxID=44475;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=Flocruz LA-130;  
RX Pubmed=15028702; DOI=10.1128/DB.186.7.2164-2172.2004;  
RA Nascimento A.L.T.O., Ko A.I., Martins E.A.L., Monteiro-Vitorello C.B.,

RA	Ho P.L., Haake D.A., Vertjovski-Almeida S., Hartskeerl R.A., Marques M.V., Oliveira M.C., Menck C.F.M., Leite L.C.C., Carrer H., Coutinho L.L., Degraeve W.M., Dellagostin O.A., El-Dorry H., Ferro E.S., Ferro M.I.T., Furlan L.R., Gambellini M., Gigliotti E.A., Goes-Neto A., Goldman G.H., Goldman M.H.S., Harakava R., Jeronimo S.M.B., Junqueira-de-Azevedo I.L.M., Kimura E.T., Kuramae E.E., Lemos J.G.M., Lemos M.V.F., Marino C.L., Nunes L.R., de Oliveira R.C., Pereira G.G., Reis M.S., Schriefer A., Siqueira W.J., Sommer P., Tsai S.M., Simpson A.J.G., Ferro J.A., Carneiro L.E.A., Kiteajima J.P., Setubal J.C., Van Sluys M.A.; "Comparative genomes of two Leptospira interrogans serovars reveals novel insights into physiology and pathogenesis.";
RT	J. Bacteriol. 186:2164-2172(2004).
RU	- FUNCTION: Part of a binding-protein-dependent transport system.
CC	- Probably responsible for the translocation of the substrate across the membrane (By similarity).
CC	- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC	- SIMILARITY: Belongs to the binding-protein-dependent transport system permease family.
DR	EMBL; AE017298; AA571503.1; -
DR	GO; GO:0016020; C:membrane; IEA.
DR	GO; GO:0005215; F:transporter activity; IEA.
DR	GO; GO:0006810; P:transport; IEA.
DR	InterPro: IPR000515; BPD transp.
DR	Pfam; PF00528; BPD_transp_1; 1.
KW	PROSITE; PS50928; ABC_TM1; 1.
KM	Complete proteome; Transmembrane; Transport.
DW	SEQUENCE 253 AA; 27677 MW; 0A145ACBE801964C CRC64;
SQ	
Query Match	52.7%; Score 59; DB 2; Length 253;
Best local Similarity	31.8%; Pred. NO. 56;
Matches	7; Conservative 11; Mismatches 4; Indels 0; Gaps 0;
Oy	1 WSHPFQEKXXXXXXXXXWSHP 22     : :::::          : :::::
Db	22 WKNPTEVPLEDSFCSVTWSP 43
RESULT 7	
OS	PRELIMINARY; PRt; 253 AA.
ID	08FRC3
AC	Q8FRC3.
DT	01-MAR-2003 (TREMblrel. 23, Created)
DT	01-MAR-2003 (TREMblrel. 23, Last sequence update)
DT	01-MAR-2004 (TREMblrel. 26, Last annotation update)
DE	Dipeptide/oligopeptide/nickel transport systems, permease components.
DE	OrderedLocNames=LA0634;
CN	Leptospira interrogans.
OS	Bacteria; Spirochaetes; Spirochaetales; Leptospiiraceae; Leptospira.
OC	NCB1_TaxId=173;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRIN56601 / Serogroup Icterohaemorrhagiae / Serovar lai;
RX	MEDLINE=22598143; PubMed=12712204; DOI=10.1038/nature01597;
RA	Ren S.-X., Fu G., Jiang X.-G., Zeng R., Miao Y.-G., Xu H., Zhang Y.-X., Xiong H., Lu G., Lu L.-F., Jiang H.-O., Jia J., Tu Y.-F., Jiang J.-X., Gu W.-Y., Zhang Y.-Q., Cai Z., Sheng H.-H., Yin H.-F., Zhang Y., Zhu G.-F., Wen M., Huang H.-L., Qian Z., Wang S.-Y., Ma W., Yao Z.-J., Shen Y., Qiang B.-O., Xia Q.-C., Guo X.-K., Danchin A., Salet Girones I., Somerville R.L., Wen Y.-M., Shi W.-H., Chen Z., Xu J.-G., Zhao G.-P.;
RA	"Unique physiological and pathogenic features of Leptospira interrogans revealed by whole-genome sequencing.";
RU	Nature 422:888-893(2003).
CC	- FUNCTION: Part of a binding-protein-dependent transport system.
CC	- Probably responsible for the translocation of the substrate across the membrane (By similarity).
CC	- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).
CC	- SIMILARITY: Belongs to the binding-protein-dependent transport system permease family.
DR	EMBL; AE011250; AAA47833.1; -
DR	GO; GO:0016021; C:integral to membrane; IEA.

```

DR GO:0005215; F:transporter activity; IEA.
DR GO:0006810; P:transport; IEA.
DR InterPro:IPR000515; BPD_transp.
DR Pfam:PF00528; BPD_transp_1; 1.
DR PROSITE:PS05928; ABC_TM1; 1.
DR Complete proteome; Transmembrane; Transport.
SQ SEQUENCE 253 AA; 27677 MW; 0A145ACBE801964C CRC64;

Query Match
Best Local Similarity 52.7%; Score 59; DB 2; Length 253;
Matches 7; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

Db 1 WSHPOFEKXXXXXXXXXWSHP 22
22 WKNPTEVFLEDFSCVTWSHP 43

RESULT 8
ID 0756P7 PRELIMINARY; PRT; 309 AA.
AC 0756P7;
DT 01-MAR-2004 (TRENBLrel. 26, Created)
DT 01-MAR-2004 (TRENBLrel. 26, Last sequence update)
DE Hypothetical protein.
GN Name=NCU04788.1;
OS Neurospora crassa.
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
OC Sordariomycetiales; Sordariales; Sordariaceae; Neurospora.
OX NCBI_TaxId=5141;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=OR74A;
RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker J.U., Read N.D.,
RA Jaffe D., Fitzhugh W., Ma L.-J., Smirnov S., Putcell S., Rehman B.,
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,
RA Seltnerikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,
RA Kolbe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,
RA Kamal M., Kamysaselis M., Mauceli E., Bielke C., Rudd S., Fishman D.,
RA Kryatova S., Kaemmen C., Metzger R.L., Perkins D.D., Kroken S.,
RA Cogoni C., Macho G., Catchside D., Li W., Pratt R.J., Osmari S.A.,
RA Desouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,
RA Naveig D.O., Alex L.A., Manhaupt G., Ebbole D.J., Freitag M.,
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.,
RA "The Genome Sequence of the Filamentous Fungus Neurospora crassa."
RT Nature 0:0-0(2003).
CC -!- CAUTION: The sequence shown here is derived from an
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is
CC preliminary data.
DR EMBL; AABX01000323; EAA31178.1; -
KW Hypothetical protein.
SQ SEQUENCE 309 AA; 32912 MW; ABE6A0CD76802891 CRC64;

Query Match
Best Local Similarity 52.7%; Score 59; DB 2; Length 309;
Matches 7; Conservative 11; Mismatches 7; Indels 0; Gaps 0;

Db 1 WSHPOFEKXXXXXXXXXWSHP 25
137 WKGPMILKAIQQLLHEVWMSHPSLD 161

RESULT 9
ID 08TK4 PRELIMINARY; PRT; 400 AA.
AC 08TK4;
DT 01-JUN-2002 (TRENBLrel. 21, Created)
DT 01-JUN-2002 (TRENBLrel. 21, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE Aldo/keto reductase.

```

```

GN OrderedLocusNames=MA0427;
OS Methanosarcina acetivorans.
OC Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC Methanosarcinaceae; Methanosarcina.
OX NCBI_TaxId=2214;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=C2A / ATCC 35395 / DSM 2834;
RX MEDLINE=21929760; PubMed=11932368; DOI=10.1101/gr.223902;
RA Galagan J.E., Nussbaum C., Roy A., Endrizzi M.G., Macdonald P.,
RA Fitzhugh W., Calvo S., Engels R., Smirnov S., Alnoor D., Brown A.,
RA Allen N., Naylor J., Strange-Thomann N., Deatellano K., Johnson R.,
RA Lincoln L., McKernan P., McKernan K., Telamas J., Titrrell A., Ye W.,
RA Zimmer A., Barber R.D., Cann I., Gramam D.E., Gramam D.A., Guss A.M.,
RA Hedderich R., Ingaram-Smith C., Kuetner H.C., Krzycki J.A.,
RA Leigh J.A., Li W., Liu J., Mukhopadhyay B., Reeve J.N., Smith K.,
RA Springer T.A., Umayan L.A., White O., White R.H., de Macario E.C.,
RA Ferry J.G., Jarrell K.F., Jing H., Macario A.J.L., Paulsen I.T.,
RA Pritchett M., Sowers K.R., Swanson R.V., Zinder S.H., Lander E.,
RA Metcalf W.W., Birren B.;
RT "The genome of Methanosarcina acetivorans reveals extensive metabolic
RT and physiological diversity."
RL Genome Res. 12:532-542(2002).
DR EMBL; AEO10703; AAM03874.1; -
DR HSSP; P52895; 1396.
DR GO:0005449; F:electron transporter activity; IEA.
DR GO:0005506; F:iron ion binding; IEA.
DR GO:0006118; P:electron transport; IEA.
DR Pfam; PF00248; Aldo_ket_red; 1.
DR Pfam; PF00037; Fer4; 1.
DR PRINTS; PR00069; ALDETRDTASE.
DR Prodom; PD000286; Aldo/ket_red; 1.
DR PROSITE; PS00198; 4F84S_FERRDOXIN; 1.
KW 4fe-4s; Complete proteome; iron; iron-bullfin; Metal-binding.
SQ SEQUENCE 400 AA; 45143 MW; 7PAD17732FEYDC29 CRC64;

Query Match
Best Local Similarity 52.7%; Score 59; DB 2; Length 400;
Matches 6; Conservative 13; Mismatches 4; Indels 0; Gaps 0;

Db 1 WSHPOFEKXXXXXXXXXWSHP 23
222 WSEADVERTPEVMALRWVWHP 244

RESULT 10
ID 09A613 PRELIMINARY; PRT; 771 AA.
AC 09A613;
DT 01-JUN-2001 (TRENBLrel. 17, Created)
DT 01-JUN-2001 (TRENBLrel. 17, Last sequence update)
DT 01-OCT-2003 (TRENBLrel. 25, Last annotation update)
DE Glucoamylase.
GN OrderedLocusNames=CC2282;
OS Caulobacter crescentus.
OC Bacteria; Proteobacteria; Alphaproteobacteria; Caulobacteriales;
OC Caulobacteriaceae; Caulobacter.
OX NCBI_TaxId=155892;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 19089 / CB15;
RX MEDLINE=21173698; PubMed=11259647; DOI=10.1073/pnas.061029298;
RA Niernan W.C., Feldblum T.V., Laub M.T., Paulsen I.T., Nelson K.E.,
RA Eisen J.A., Heidelberg J.F., Alley M.R.K., Ohta N., Maddock J.R.,
RA Potocka I., Nelson W.C., Newton A.S., Gwin M.L., Haft D.H., Ely B.,
RA Deboy R.T., Dodson R.U., Durkin A.S., Khouri H.M., Shetty J.,
RA Berry K.J., Uteerback T.R., Tran K., Wolf A.M., Vamathevan J.J.,
RA Ermolaeva M.D., White O., Salzberg S.L., Venter J.C., Shapiro L.,
RA Fraser C.M.;
RT "Complete genome sequence of Caulobacter crescentus."
RT Proc. Natl. Acad. Sci. U.S.A. 98:4136-4141(2001).
DR EMBL; AEO05898; AAK24253.1; -

```



DT 25-OCT-2004 (TReMBLrel. 28, Created)  
DT 25-OCT-2004 (TReMBLrel. 28, Last sequence update)  
DT 25-OCT-2004 (TReMBLrel. 28, Last annotation update)  
DE Hypothetical protein OSJNB0017118.11 (Hypothetical protein  
DE OSJNB0095104.17).  
CN Name=OSJNB0017118.11; Synonym=OSJNB0095104.17;  
OS Oryza sativa (japonica cultivar-group). Embryophyta; Tracheophyta;  
OC Eukaryota; Viridiplantae; Streptophyta; Eudicotyledons; Poales;  
OC Spermatophyta; Magnoliophyta; Liliopsida; Poaceae;  
OC Eriocarpaceae; Oryzae; Oryza.  
OX NCBI\_TaxID=39947;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Sasaki T., Matsunoto T., Katayose Y.,  
RL Submitted (SEP-2002) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AP005724; BAD2919.1; -.  
DR EMBL; AP005701; BAD2913.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 202 AA; 22026 MW; 23392FDF138284FB CRC64;

Query Match 50.9%; Score 57; DB 2; Length 202;  
Best Local Similarity 27.3%; Pred. No. 81;  
Matches 6; Conservative 12; Mismatches 4; Indels 0; Gaps 0;

OY 1 WSHPOFEKXXXXXXXXXXKSHP 22  
DB 148 WGHPSFSGRCRCPERRRDP 169

RESULT 15  
ID 07SAB0 PRELIMINARY; PRT; 1111 AA.  
AC 07SAB0;  
DT 01-MAR-2004 (TReMBLrel. 26, Created)  
DT 01-MAR-2004 (TReMBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TReMBLrel. 26, Last annotation update)  
DE Predicted protein.  
GN Name=NCU06321.1;  
OS Neurospora crassa.  
OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.  
OX NCBI\_TaxID=5141;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=OR74A;  
RA Galagan J.E., Borkovich K.A., Selker E.U., Read N.D.,  
RA Jaffe D., Fitzhugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,  
RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,  
RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,  
RA Selkrenikoff C.P., Kinsey J.A., Braun E.L., Zeller A., Schulte U.,  
RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,  
RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,  
RA Kamal M., Kamysaselis M., Mauceli E., Bielek C., Rudd S., Fishman D.,  
RA Kryzofova S., Rasmussen C., Metzger R.L., Perkins D.D., Kroken S.,  
RA Cogoni C., Macino G., Catcheside D., Li W., Pratt R.J., Omani S.A.,  
RA Desouza C.C., Glas L., Orbach M.J., Berglund J., Voelker R.,  
RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,  
RA Natvig D.O., Alex L.A., Manhaupt G., Ebbole D.O., Freitag M.,  
RA Paulsen I., Sachs M.S., Lander E.S., Nussbaum C., Birren B.,  
RT "The Genome Sequence of the Filamentous Fungus Neurospora crassa."  
RL Nature 0:0-0 (2003).  
CC -! CAUTION: The sequence shown here is derived from an  
CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
CC preliminary data.  
CC DR EMBL; AABX0100205; EAA33347.1; -.  
SQ SEQUENCE 1111 AA; 121374 MW; 31264510AF33B363 CRC64;

Query Match 50.9%; Score 57; DB 2; Length 1111;  
Best Local Similarity 28.0%; Pred. No. 5.1e+02;  
Matches 7; Conservative 10; Mismatches 8; Indels 0; Gaps 0;

OY 1 WSHPOFEKXXXXXXXXXXKSHP 25  
DB 148 WGHPSFSGRCRCPERRRDP 169

DB 292 WGHDDVVEWQHDPVVEWQHDPDE 316

Search completed: March 2, 2005, 12:44:29  
Job time : 97.6016 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 117.106 Seconds  
(without alignments)  
85.869 Million cell updates/sec

Title: SEQ14  
Perfect score: 112  
Sequence: 1 wshpqfexkxxxxxxxxxwshpqfek 26

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues  
Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16dec04:\*  
1: geneseqp19808:\*  
2: geneseqp19908:\*  
3: geneseqp20008:\*  
4: geneseqp20018:\*  
5: geneseqp20028:\*  
6: geneseqp20038:\*  
7: geneseqp20048:\*  
8: geneseqp20058:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99	88.4	24	6	ABP60362
2	97.5	87.1	35	6	ABP60369
3	97	86.6	36	6	ABP60370
4	72	64.3	236	8	AD019051
5	70	64.3	236	8	AD019053
6	70	62.5	24	6	ABP60363
7	68	60.7	209	6	ADA00709
8	68	60.7	258	6	ADA27293
9	68	60.7	258	6	ADA00702
10	68	60.7	400	6	ADA27291
11	68	60.7	400	6	ADA27292
12	68	60.7	400	6	ADA00700
13	68	60.7	400	6	ADA00707
14	68	60.7	400	6	ADA00701
15	66	58.9	646	8	ADS20251
16	66	58.9	858	8	ADS20250
17	65	58.0	539	6	ABG74883
18	65	58.0	539	6	AD139158
19	64.5	57.6	485	8	ADS30249
20	64.5	57.6	697	8	ADS30227
21	64	57.1	699	7	ADJ58628
22	64	57.1	699	8	ADQ18833
23	64	57.1	699	8	ADQ39188
24	64	57.1	699	8	ADQ39189
25	63	56.2	183	6	ADA27294

26	63	56.2	183	6	ADA00703	AdA00703 Modified
27	63	56.2	467	8	ADP18461	Adp18461 Arthrobac
28	62	55.4	56	4	AAE04104	AAe04104 Human gen
29	62	55.4	56	4	AAE04132	AAe04132 Human gen
30	62	55.4	56	5	ABG64354	ABg64354 Human alb
31	62	55.4	56	5	ABG64356	ABg64356 Human alb
32	62	55.4	56	6	ADA98358	Ada98358 Human sec
33	62	55.4	56	6	ADA44167	Ada44167 Human sec
34	62	55.4	56	7	ADC20522	Adc20522 Human sec
35	62	55.4	56	7	ADL10783	Adl10783 Human sec
36	62	55.4	56	8	ADL77621	Adl77621 Albumin f
37	62	55.4	56	8	ADL77623	Adl77623 Albumin f
38	62	55.4	57	4	AAU01630	AAu01630 Human sec
39	62	55.4	267	8	ADR70320	Adr70320 Polioviru
40	62	55.4	268	8	ADR70385	Adr70385 Polioviru
41	62	55.4	268	8	ADR70319	Adr70319 Polioviru
42	62	55.4	268	8	ADR70386	Adr70386 Polioviru
43	62	55.4	269	8	ADR28054	Adr28054 NPb polyP
44	62	55.4	269	8	ADS17525	Ads17525 Amino aci
45	62	55.4	273	8	ADO25157	Ado25157 Melanoma

## ALIGNMENTS

RESULT 1	ABP60362	standard; peptide; 24 AA.
XX	AC	ABP60362;
XX	DT	28-MAR-2003 (first entry)
XX	DE	Streptavidin binding peptide SEQ ID NO 3.
XX	KW	Streptavidin; protein chip; microtitre plate; detection.
XX	OS	Synthetic.
XX	FH	Key
XX	FT	Misc-difference 9..16
XX	FT	Location/Qualifiers
XX	XX	label= unknown
XX	PN	DE1013776-A1.
XX	PD	02-OCT-2002.
XX	PF	21-MAR-2001; 2001DE-01013776.
XX	PR	21-MAR-2001; 2001DE-01013776.
XX	PA	(BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX	PI	Schmidt T;
XX	DR	WPI; 2003-031166/03.
XX	PT	New isolated peptide; useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.
XX	PS	Disclosure; Page 4; 18pp; German.
XX	CC	The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Gln, Asn or Met. (II), which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides stronger binding than a single tag, but are displaced by a competitor. (I) does not interfere with the function of attached proteins (II) (so it may not be essential to remove it); facilitates detection and has easily

CC controllable binding properties. (I) is particularly used for purifying  
CC FP from dilute solution in batch formats (which use simpler apparatus  
CC than column methods and result in lower loss of FP). The present sequence  
CC is that of a streptavidin binding peptide disclosed with the invention  
XX  
SQ Sequence 24 AA;

Query Match 88.4%; Score 99; DB 6; Length 24;  
Best Local Similarity 92.3%; Pred. No. 9,7e-06;  
Matches 24; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

QY 1 WSHPOFEKXXXXXXXXXXWSHPQFEK 26  
Db 1 WSHPOFEK-XXXXXXXXXXWSHPQFEK 24

## RESULT 2

ABP60369 ID ABP60369 standard; peptide; 35 AA.

AC ABP60369;

DT 28-MAR-2003 (first entry)

DE Streptavidin binding peptide SEQ ID NO 10.

KW Streptavidin; protein chip; microtitre plate; detection.

OS Synthetic.

FT Key Location/Qualifiers

FT Misc-difference 9..27 /label= unknown

FT /note= "optionally deleted for 1-15 residues"

PN DE10113776-A1.

XX 02-OCT-2002.

PD 21-MAR-2001; 2001DE-01013776.

PF 21-MAR-2001; 2001DE-01013776.

PR (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

PA Schmidt T;

PI WPI; 2003-031166/03.

DR New isolated peptide, useful as affinity purification tag for recombinant

PT protein, comprises at least two high-affinity streptavidin-binding

PS modules.

XX Claim 7; Page 16; 18pp; German.

CC The invention relates to an isolated peptide (I) comprising at least two  
CC individual modules separated by 0-50 amino acids, with each containing at  
CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain  
CC streptavidin binding modules, are useful as affinity handles for  
CC purification of recombinant fusion proteins (FP), also for detecting FP,  
CC e.g. on protein chips or microtitre plates. The modules in (I) bind  
CC strongly to streptavidin, with a co-operative effect that provides  
CC stronger binding than a single tag, but are displaced by a competitor.  
CC (I) does not interfere with the function of attached proteins (II) (so it  
CC may not be essential to remove it); facilitates detection and has easily  
CC controllable binding properties. (I) is particularly used for purifying  
CC FP from dilute solution in batch formats (which use simpler apparatus  
CC than column methods and result in lower loss of FP). The present sequence  
CC is that of a streptavidin binding peptide disclosed with the invention  
XX

SQ Sequence 35 AA;

Query Match 87.1%; Score 97.5; DB 6; Length 35;

Best Local Similarity 74.3%; Pred. No. 2.6e-05;  
Matches 26; Conservative 0; Mismatches 0; Indels 9; Gaps 1;

QY 1 WSHPOFEK-----XXXXXXXXXXWSHPQFEK 26  
Db 1 WSHPOFEKXXXXXXXXXXWSHPQFEK 35

## RESULT 3

ABP60370 ID ABP60370 standard; peptide; 36 AA.

AC ABP60370;

DT 28-MAR-2003 (first entry)

DE Streptavidin binding peptide SEQ ID NO 11.

KW Streptavidin; protein chip; microtitre plate; detection.

OS Synthetic.

FT Key Location/Qualifiers

FT Region 9..28 /note= "GGGS repeats 2-5 optionally absent, residues 13-

FT 28" /label= GGGS\_repeat

PN DE10113776-A1.

XX 02-OCT-2002.

PD 21-MAR-2001; 2001DE-01013776.

PF 21-MAR-2001; 2001DE-01013776.

PR (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

PA Schmidt T;

PI WPI; 2003-031166/03.

DR New isolated peptide, useful as affinity purification tag for recombinant

PT protein, comprises at least two high-affinity streptavidin-binding

PS modules.

XX Claim 8; Page 16; 18pp; German.

CC The invention relates to an isolated peptide (I) comprising at least two  
CC individual modules separated by 0-50 amino acids, with each containing at  
CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain  
CC streptavidin binding modules, are useful as affinity handles for  
CC purification of recombinant fusion proteins (FP), also for detecting FP,  
CC e.g. on protein chips or microtitre plates. The modules in (I) bind  
CC strongly to streptavidin, with a co-operative effect that provides  
CC stronger binding than a single tag, but are displaced by a competitor.  
CC (I) does not interfere with the function of attached proteins (II) (so it  
CC may not be essential to remove it); facilitates detection and has easily  
CC controllable binding properties. (I) is particularly used for purifying  
CC FP from dilute solution in batch formats (which use simpler apparatus  
CC than column methods and result in lower loss of FP). The present sequence  
CC is that of a streptavidin binding peptide disclosed with the invention  
XX

SQ Sequence 36 AA;

Query Match 86.6%; Score 97; DB 6; Length 36;  
Best Local Similarity 44.4%; Pred. No. 3.2e-05;  
Matches 16; Conservative 10; Mismatches 0; Indels 10; Gaps 1;

QY 1 WSHPOFEK-----XXXXXXXXXXWSHPQFEK 26  
Db 1 WSHPOFEKGGSGGSGGSGGSGGSGGSGWSHPQFEK 36



```

RESULT 4
AD019051 standard; protein; 236 AA.
XX
XX
AC AD019051;
XX
DT 26-AUG-2004 (first entry)
XX
DE Murine antibody M1 Fab fragment heavy chain.
XX
KM Mouse; antibody; M1 Fab; heavy chain; Fab; scFv; abused drug; morphine;
KM THC; amphetamine; environmental hazard; toxic compound;
KM microbial process; metabolic process; drug monitoring;
KM pharmacological research.
XX
OS Mus sp.
XX
XX WO2004046733-A1.
XX
XX 03-JUN-2004.
XX
XX 17-NOV-2003; 2003WO-FI000875.
XX
XX 18-NOV-2002; 2002FI-00002048.
XX
XX (VALM ) VALTION TEKNIILLINEN TUTKIMUSKESKUS.
XX
XX Pulli T, Hoeyhtyae M, Takkinen K, Soederlund H;
XX
XX WPI; 2004-420710/39.
XX
PT Non-competitive immunoassay for small analyte, useful for assaying drug
PT of abuse (e.g., morphine), comprises reacting a sample of analyte with a
PT reagent pair comprising a first binding partner and a second binding
PT partner.
XX
XX Claim 21; SEQ ID NO 2; 35pp; English.
XX
CC The invention relates to a non-competitive immunoassay for a small
CC analyte, comprising reacting a sample containing the analyte with a
CC reagent pair comprising a first binding partner that binds to the analyte
CC and a second binding partner that binds to the complex of the analyte
CC and the first binding partner, and determining the binding of the second
CC binding partner, thus indicating the presence of the analyte in the
CC sample. The first and second binding partners are antibody fragments Fab
CC or scFv. The reagent pair is useful in a non-competitive immunoassay for
CC a small analyte, particularly for assaying drugs of abuse e.g., morphine,
CC THC or amphetamine. The immunoassay is useful for detecting environmental
CC hazards, toxic compounds in food and feed, chemicals indicative of
CC ongoing processes (e.g., microbial processes in buildings, metabolic
CC processes of living organisms) and in clinical tests, drug monitoring and
CC pharmacological research. This sequence represents the murine antibody M1
CC Fab fragment heavy chain, used in the method of the invention.
XX
SQ Sequence 236 AA;
XX
Query Match 64.3%; Score 72; DB 8; Length 236;
Best Local Similarity 37.9%; Pred. No. 1.1;
Matches 11; Conservative 13; Mismatches 1; Indels 4; Gaps 1;
QY 2 SHP-----QFEKXXXXXXXXXXWSHPQFEK 26
DB 208 AHPASTKVDKIVPRDCGTSMHPQFEK 236
XX
RESULT 5
AD019053
ID AD019053 standard; protein; 236 AA.
XX
XX AD019053;
XX

```

```

DT 26-AUG-2004 (first entry)
XX
XX Murine antibody M2 Fab fragment heavy chain.
XX
DE Mouse; antibody; M2 Fab; heavy chain; Fab; scFv; abused drug; morphine;
XX
KM THC; amphetamine; environmental hazard; toxic compound;
KM microbial process; metabolic process; drug monitoring;
KM pharmacological research.
XX
OS Mus sp.
XX
XX WO2004046733-A1.
XX
XX 03-JUN-2004.
XX
XX 17-NOV-2003; 2003WO-FI000875.
XX
XX 18-NOV-2002; 2002FI-00002048.
XX
XX (VALM ) VALTION TEKNIILLINEN TUTKIMUSKESKUS.
XX
XX Pulli T, Hoeyhtyae M, Takkinen K, Soederlund H;
XX
XX WPI; 2004-420710/39.
XX
XX
PT Non-competitive immunoassay for small analyte, useful for assaying drug
PT of abuse (e.g., morphine), comprises reacting a sample of analyte with a
PT reagent pair comprising a first binding partner and a second binding
PT partner.
XX
XX Claim 21; SEQ ID NO 4; 35pp; English.
XX
CC The invention relates to a non-competitive immunoassay for a small
CC analyte, comprising reacting a sample containing the analyte with a
CC reagent pair comprising a first binding partner that binds to the analyte
CC and a second binding partner that binds to the complex of the analyte and
CC the first binding partner, and determining the binding of the second
CC binding partner, thus indicating the presence of the analyte in the
CC sample. The first and second binding partners are antibody fragments Fab
CC or scFv. The reagent pair is useful in a non-competitive immunoassay for
CC a small analyte, particularly for assaying drugs of abuse e.g., morphine,
CC THC or amphetamine. The immunoassay is useful for detecting environmental
CC hazards, toxic compounds in food and feed, chemicals indicative of
CC ongoing processes (e.g., microbial processes in buildings, metabolic
CC processes of living organisms) and in clinical tests, drug monitoring and
CC pharmacological research. This sequence represents the murine antibody M2
CC Fab fragment heavy chain, used in the method of the invention.
XX
SQ Sequence 236 AA;
XX
Query Match 64.3%; Score 72; DB 8; Length 236;
Best Local Similarity 37.9%; Pred. No. 1.1;
Matches 11; Conservative 13; Mismatches 1; Indels 4; Gaps 1;
QY 2 SHP-----QFEKXXXXXXXXXXWSHPQFEK 26
DB 208 AHPASTKVDKIVPRDCGTSMHPQFEK 236
XX
RESULT 6
ABP60363
ID ABP60363 standard; peptide; 24 AA.
XX
XX ABP60363;
XX
XX
DT 28-MAR-2003 (first entry)
XX
DE Streptavidin binding peptide SEQ ID NO 4.
XX
XX Streptavidin; protein chip; microtitre plate; detection.
XX
XX Synthetic.
XX

```

```

FH Key Location/Qualifiers
FT Misc-difference 9..18
FT /label= unknown
FT Misc-difference 22..24
FT /label= unknown
FT DE10113776-A1.
XX
XX 02-OCT-2002.
XX
XX 21-MAR-2001; 2001DE-01013776.
XX
XX 21-MAR-2001; 2001DE-01013776.
XX
XX (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
XX Schmidt T;
XX
XX WPI; 2003-031166/03.
XX
XX
XX New isolated peptide, useful as affinity purification tag for recombinant
XX protein, comprises at least two high-affinity streptavidin-binding
XX modules.
XX
XX Disclosure; Page 4; 18pp; German.
XX
XX The invention relates to an isolated peptide (I) comprising at least two
XX individual modules separated by 0-50 amino acids, with each containing at
XX least one motif His-Pro-X where X = Gln, Asn or Met. (1), which contain
XX streptavidin binding modules, are useful as affinity handles for
XX purification of recombinant fusion proteins (FP), also for detecting FP,
XX e.g. on protein chips or microtitre plates. The modules in (I) bind
XX strongly to streptavidin, with a co-operative effect that provides
XX stronger binding than a single tag, but are displaced by a competitor.
XX (I) does not interfere with the function of attached proteins (II) (so it
XX may not be essential to remove it); facilitates detection and has easily
XX controllable binding properties. (I) is particularly used for purifying
XX FP from dilute solution in batch formats (which use simpler apparatus
XX than column methods and result in lower loss of FP). The present sequence
XX is that of a streptavidin binding peptide disclosed with the invention
XX
XX Sequence 24 AA;
XX
XX Query Match 62.5%; Score 70; DB 6; Length 24;
XX Best Local Similarity 91.3%; Pred. No. 0.084;
XX Matches 21; Conservative 0; Mismatches 0; Indels 2; Gaps 1;
XX
XX 1 WSHPOFEKXXXXXXXXXXWSHPQ 23
XX 1 WSHPOFEKXXXXXXXXXX--HPQ 21
XX
XX
XX RESULT 7
XX ADA00709
XX ID ADA00709 standard; protein; 209 AA.
XX
XX AC ADA00709;
XX
XX 06-NOV-2003 (first entry)
XX
XX DE Modified hNGAL related fusion protein SEQ ID NO:22.
XX
XX KW mutein; human neutrophil gelatinase-associated lipocalin; hNGAL;
XX rat alpha2-microglobulin-related protein; A2m; mouse 24p3/uterocalin;
XX 24p3; mutagenesis; fusion protein; cytostatic; gene therapy;
XX tumour imaging; cancer therapy.
XX
XX OS Synthetic.
XX
XX OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX FH 1..21
XX FT Peptide /label= signal

```

```

FT Protein 22..209
FT /note= "modified hNGAL and strep-tag II fusion protein"
FT Region 22..199
FT /note= "mature hNGAL"
FT Region 200..209
FT /note= "strep-tag II affinity tag"
XX
XX WO2003029463-A2.
XX
XX 10-APR-2003.
XX
XX 18-SEP-2002; 2002MO-EP010490.
XX
XX 27-SEP-2001; 2001MO-EP011213.
XX 16-APR-2002; 2002MO-EP004223.
XX
XX (PIER-) PIERIS PROTEOLAB AG.
XX
XX Skerra A, Schlehuber S;
XX
XX WPI; 2003-372000/35.
XX
XX N-PSDB; ADA00708.
XX
XX Generating a mutein of a protein, e.g. human neutrophil gelatinase-
XX associated lipocalin, rat alpha2-microglobulin-related protein or mouse
XX 24p3/uterocalin, comprises subjecting the protein to mutagenesis.
XX
XX Disclosure; Page 107-108; 122pp; English.
XX
XX The present invention describes a method for generating a mutein of a
XX protein selected from a human neutrophil gelatinase-associated lipocalin
XX (hNGAL), rat alpha2-microglobulin-related protein (A2m) and a mouse
XX 24p3/uterocalin (24p3), where the mutein has a detectable affinity to a
XX given target, comprising subjecting the protein to mutagenesis at
XX positions 33-54, 66-83, 94-106 and/or 123-136 of hNGAL, resulting in
XX CC mutants of the protein. Also described: (1) a mutein of hNGAL, A2m or
XX 24p3 having detectable binding affinity to a given target, obtained by
XX the method described above; (2) a fusion protein comprising the mutein of
XX hNGAL, A2m or 24p3, where an enzyme, a protein or a protein domain, a
XX peptide, a signal sequence and/or an affinity tag is operably fused to
XX the amino or carboxy terminus of the mutein; (3) a nucleic acid molecule
XX comprising a sequence encoding the mutein of hNGAL, A2m or 24p3 or the
XX fusion protein of (2); and (4) a pharmaceutical composition comprising
XX the mutein of hNGAL, A2m or 24p3 or the fusion protein described above,
XX and a carrier. The muteins have cytostatic activity, and can be used in
XX gene therapy. The method is useful in generating or producing a mutein of
XX hNGAL, A2m or 24p3 or a fusion protein. The mutein of hNGAL, A2m or 24p3
XX or the fusion protein is useful in detecting a given target by contacting
XX the mutein with a sample suspected of containing the given target under
XX conditions that allow complex formation between the mutein and the given
XX target, and determining the complexed mutein by a suitable signal. The
XX given target is a protein or protein domain, a peptide, a nucleic acid
XX molecule, an organic molecule or a metal complex and the detection is
XX carried out for validation of the protein as a pharmacological drug
XX target. The mutein may also be used in medicine, e.g. for tumour imaging
XX or directly for cancer therapy. The present sequence represents a
XX modified hNGAL and Strep-tag II fusion protein given in the
XX exemplification of the present invention.
XX
XX Sequence 209 AA;
XX
XX Query Match 60.7%; Score 68; DB 6; Length 209;
XX Best Local Similarity 40.0%; Pred. No. 3.4;
XX Matches 10; Conservative 11; Mismatches 4; Indels 0; Gaps 0;
XX
XX 2 SHPOFEKXXXXXXXXXXWSHPQFEK 26
XX 1 NHTVFPVPIDQCIDGSAMSHPOFEK 209
XX
XX
XX RESULT 8
XX ADA27293
XX ID ADA27293 standard; protein; 258 AA.

```





PT	24p3/uterocalin,	comprises subjecting the protein to mutagenesis.
XX		
PS	Example 3;	Page 94-95; 122pp; English.
XX		
CC	The present invention describes a method for generating a mutein of a	
CC	protein selected from a human neutrophil gelatinase-associated lipocalin	
CC	(hNGAL), rat alpha2-microglobulin-related protein (A2m) and a mouse	
CC	24p3/uterocalin (24p3), where the mutein has a detectable affinity to a	
CC	given target, comprising subjecting the protein to mutagenesis at	
CC	positions 33-54, 66-83, 94-106 and/or 123-136 of hNGAL, resulting in	
CC	muteins of the protein. Also described: (1) a mutein of hNGAL, A2m or	
CC	24p3 having detectable binding affinity to a given target, obtained by	
CC	the method described above; (2) a fusion protein comprising the mutein of	
CC	hNGAL, A2m or 24p3, where an enzyme, a protein or a protein domain, a	
CC	peptide, a signal sequence and/or an affinity tag is operably fused to	
CC	the amino or carboxy terminus of the mutein; (3) a nucleic acid molecule	
CC	comprising a sequence encoding the mutein of hNGAL, A2m or 24p3 or the	
CC	fusion protein of (2); and (4) a pharmaceutical composition comprising	
CC	the mutein of hNGAL, A2m or 24p3 or the fusion protein described above,	
CC	and a carrier. The muteins have cytostatic activity, and can be used in	
CC	gene therapy. The method is useful in generating or producing a mutein of	
CC	hNGAL, A2m or 24p3 or a fusion protein. The mutein of hNGAL, A2m or 24p3	
CC	or the fusion protein is useful in detecting a given target by contacting	
CC	the mutein with a sample suspected of containing the given target under	
CC	conditions that allow complex formation between the mutein and the given	
CC	target, and determining the complexed mutein by a suitable signal. The	
CC	given target is a protein or protein domain, a peptide, a nucleic acid	
CC	molecule, an organic molecule or a metal complex and the detection is	
CC	carried out for validation of the protein as a pharmacological drug	
CC	target. The mutein may also be used in medicine, e.g. for tumour imaging	
CC	or directly for cancer therapy. The present sequence represents a	
CC	modified hNGAL, Strep-tag II and phage coat protein pIII fragment fusion	
CC	protein given in the exemplification of the present invention.	
XX		
SQ	Sequence 400 AA:	
Query Match	60.7%; Score 68;	DB 6; Length 400;
Best Local Similarity	40.0%; Pred. No. 8.4;	
Matches 10;	Conservative * 11;	Mismatches 4; Indels 0; Gaps 0;
CY	2 SHPOFEKXXXXXXXXXMSHPQEFK 26	
	:  :   : : : : :	
Dd	185 NHIVFPVPIDQCIDGSAMSHPOFEK 209	
RESULT 13		
ADA00707		
ID	ADA00707 standard;	protein; 400 AA.
XX		
AC	ADA00707;	
XX		
DT	06-NOV-2003	(first entry)
DE	Modified hNGAL related fusion protein SEQ ID NO:20.	
XX		
KW	mutcin; human neutrophil gelatinase-associated lipocalin; hNGAL;	
KW	rat alpha2-microglobulin-related protein; A2m; mouse 24p3/uterocalin;	
KW	24p3; mutagenesis; fusion protein; cytostatic; gene therapy;	
KM	tumour imaging; cancer therapy.	
XX		
OS	Synthetic.	
OS	Homo sapiens.	
XX		
FH	Key	Location/Qualifiers
FT	Peptide	1..21
FT		/label= signal
FT	Protein	22..400
FT		/note= "modified hNGAL, Strep-tag II and phage coat
FT	Region	protein pIII fragment fusion protein"
FT		22..199
FT		/note= "mature hNGAL"
FT	Region	200..209
FT		/note= "Strep-tag II affinity tag"

```

FT Misc-difference 210
FT /note= "amber stop codon"
FT Region 211..400
FT /note= "coat protein pIII fragment 217-406"
XX
XX WO2003029463-A2.
XX
XX 10-APR-2003.
XX
XX 18-SEP-2002; 2002WO-EP010490.
XX
XX 27-SEP-2001; 2001WO-EP011213.
XX
XX 16-APR-2002; 2002WO-EP004223.
XX
XX (PIER-) PIERIS PROTEOLAB AG.
XX
XX Skerra A, Schlhuber S,
XX
XX WPI; 2003-372000/35.
XX
XX N-PSDB; ADA00706.
XX
XX
XX Generating a muttein of a protein, e.g. human neutrophil gelatinase-
XX associated lipocalin, rat alpha2-microglobulin-related protein or mouse
XX 24p3/uterocalin, comprises subjecting the protein to mutagenesis.
XX
XX Disclosure; Page 104-105; 122pp; English.
XX
XX The present invention describes a method for generating a muttein of a
XX protein selected from a human neutrophil gelatinase-associated lipocalin
XX (hNGL), rat alpha2-microglobulin-related protein (A2m) and a mouse
XX 24p3/uterocalin (24p3), where the muttein has a detectable affinity to a
XX given target, comprising subjecting the protein to mutagenesis at
XX positions 33-54, 66-83, 94-106 and/or 123-136 of hNGL, resulting in
XX mutteins of the protein. Also described: (1) a muttein of hNGL, A2m or
XX 24p3 having detectable binding affinity to a given target, obtained by
XX the method described above; (2) a fusion protein comprising the muttein of
XX hNGL, A2m or 24p3, where an enzyme, a protein or a protein domain, a
XX peptide, a signal sequence and/or an affinity tag is operably fused to
XX the amino or carboxy terminus of the muttein; (3) a nucleic acid molecule
XX comprising a sequence encoding the muttein of hNGL, A2m or 24p3 or the
XX fusion protein of (2); and (4) a pharmaceutical composition comprising
XX the muttein of hNGL, A2m or 24p3 or the fusion protein described above,
XX and a carrier. The mutteins have cytostatic activity, and can be used in
XX gene therapy. The method is useful in generating or producing a muttein of
XX hNGL, A2m or 24p3 or a fusion protein. The muttein of hNGL, A2m or 24p3
XX or the fusion protein is useful in detecting a given target by contacting
XX the muttein with a sample suspected of containing the given target under
XX conditions that allow complex formation between the muttein and the given
XX target, and determining the complexed muttein by a suitable signal. The
XX molecule, an organic molecule or a metal complex and the detection is
XX carried out for validation of the protein as a pharmacological drug
XX target. The muttein may also be used in medicine, e.g. for tumour imaging
XX or directly for cancer therapy. The present sequence represents a
XX modified hNGL, Strep-tag II and phage coat protein pIII fragment fusion
XX protein given in the exemplification of the present invention.
XX
XX Sequence 400 AA;
XX
XX Query Match 60.7%; Score 68; DB 6; Length 400;
XX Best Local Similarity 40.0%; Pred. No. 8.4;
XX Matches 10; Conservative 11; Mismatches 4; Indels 0; Gaps 0;
XX
XX 2 SHPOFEKXXXXXXXXXXWSHPQFEK 26
XX :|||:|||||:|||||
XX 185 NHIVFVPIDQICDSAMSHPOFEK 209
XX
XX RESULT 14
XX ID ADA00701 standard; protein; 400 AA.
XX AC ADA00701;

```

```

XX
XX 06-NOV-2003 (first entry)
XX
XX Modified hNGL related fusion protein SEQ ID NO:14.
XX
XX muttein; human neutrophil gelatinase-associated lipocalin; hNGL;
XX rat alpha2-microglobulin-related protein; A2m; mouse 24p3/uterocalin;
XX 24p3; mutagenesis; fusion protein; cytostatic; gene therapy;
XX tumour imaging; cancer therapy.
XX
XX Synthetic.
XX OS Homo sapiens.
XX
XX Key Location/Qualifiers
XX Peptide 1..21
XX Protein 22..400
XX /note= "modified hNGL, Strep-tag II and phage coat
XX protein pIII fragment fusion protein"
XX
XX Region 22..199
XX /note= "mature hNGL"
XX
XX Region 200..209
XX /note= "Strep-tag II affinity tag"
XX
XX Misc-difference 210
XX /note= "amber stop codon"
XX
XX Region 211..400
XX /note= "coat protein pIII fragment 217-406"
XX
XX WO2003029463-A2.
XX
XX 10-APR-2003.
XX
XX 18-SEP-2002; 2002WO-EP010490.
XX
XX 27-SEP-2001; 2001WO-EP011213.
XX
XX 16-APR-2002; 2002WO-EP004223.
XX
XX (PIER-) PIERIS PROTEOLAB AG.
XX
XX Skerra A, Schlhuber S,
XX
XX WPI; 2003-372000/35.
XX
XX N-PSDB; ADA00732.
XX
XX Generating a muttein of a protein, e.g. human neutrophil gelatinase-
XX associated lipocalin, rat alpha2-microglobulin-related protein or mouse
XX 24p3/uterocalin, comprises subjecting the protein to mutagenesis.
XX
XX Disclosure; Page 96-98; 122pp; English.
XX
XX The present invention describes a method for generating a muttein of a
XX protein selected from a human neutrophil gelatinase-associated lipocalin
XX (hNGL), rat alpha2-microglobulin-related protein (A2m) and a mouse
XX 24p3/uterocalin (24p3), where the muttein has a detectable affinity to a
XX given target, comprising subjecting the protein to mutagenesis at
XX positions 33-54, 66-83, 94-106 and/or 123-136 of hNGL, resulting in
XX mutteins of the protein. Also described: (1) a muttein of hNGL, A2m or
XX 24p3 having detectable binding affinity to a given target, obtained by
XX the method described above; (2) a fusion protein comprising the muttein of
XX hNGL, A2m or 24p3, where an enzyme, a protein or a protein domain, a
XX peptide, a signal sequence and/or an affinity tag is operably fused to
XX the amino or carboxy terminus of the muttein; (3) a nucleic acid molecule
XX comprising a sequence encoding the muttein of hNGL, A2m or 24p3 or the
XX fusion protein of (2); and (4) a pharmaceutical composition comprising
XX the muttein of hNGL, A2m or 24p3 or the fusion protein described above,
XX and a carrier. The mutteins have cytostatic activity, and can be used in
XX gene therapy. The method is useful in generating or producing a muttein of
XX hNGL, A2m or 24p3 or a fusion protein. The muttein of hNGL, A2m or 24p3
XX or the fusion protein is useful in detecting a given target by contacting
XX the muttein with a sample suspected of containing the given target under
XX conditions that allow complex formation between the muttein and the given
XX target, and determining the complexed muttein by a suitable signal. The
XX given target is a protein or protein domain, a peptide, a nucleic acid

```

CC molecule, an organic molecule or a metal complex and the detection is  
 CC carried out for validation of the protein as a pharmacological drug  
 CC target. The molecule may also be used in medicine, e.g. for tumour imaging  
 CC or directly for cancer therapy. The present sequence represents a  
 CC modified hNGAL, Strep-tag II and phage coat protein pIII fragment fusion  
 CC protein given in the exemplification of the present invention.

XX Sequence 400 AA;

Query Match 60.7%; Score 68; DB 6; Length 400;

Best Local Similarity 40.0%; Pred. No. 8.4;  
 Matches 10; Conservative 11; Mismatches 4; Indels 0; Gaps 0;

Qy 2 SHPOEKXXXXXXXXXXSHPOFEK 26

Db 185 NHIVPPVPIDCIDGSAMSHPOFEK 209

RESULT 15

ADS20251  
 ID ADS20251 standard; protein, 646 AA.

AC ADS20251;

DT 18-NOV-2004 (first entry)

DE Human aggrecanase modified ADAMTS4 (mTS4) protein - SEQ ID 49.

XX ADAMTS4; a disintegrin-like and metalloprotease;

KW thrombospondin type 1 motif 4; reprotysin; zinc metalloprotease;  
 KW aggrecanase; osteopontin; anti-inflammatory; antirheumatic;  
 KW cystostatic; osteoarthritis; glioma; cancer; inflammatory joint;  
 KW rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;  
 KW human; chromosome 1q21-q23; enzyme.

XX Home sapiens.

OS Synthetic.

XX WO2004011637-A2.

XX 05-FEB-2004.

XX 29-JUL-2003; 2003WO-US023484.

XX 29-JUL-2002; 2002US-0398721P.

XX (AMHP ) WYETH.

XX (CORC/) CORCORAN C J.

XX (FLAN/) FLANNERY C R.

XX (ZENG/) ZENG W.

XX (RACI/) RACIE L A.

XX (MCDON/) MCDONAGH T.

XX (FREE/) FREEMAN B A.

XX (GEOR/) GEORGIADIS K E.

XX (LAVA/) LAVALLIE E R.

XX Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T;

XX Freeman BA, Georgiadis KE, Lavallie ER;

XX WPI; 2004-143660/14.

XX Claim 9; SEQ ID NO 49; 117pp; English.

XX The invention relates to a novel isolated, modified ADAMTS4 (a

XX disintegrin-like and metalloprotease (reprotysin type) with

XX thrombospondin type 1 motif 4) protein with improved stability compared

XX to a naturally occurring, full-length ADAMTS4 protein, where the modified

XX protein differs from the naturally-occurring, full-length ADAMTS4 protein

XX by at least one amino acid. ADAMTS proteins are a subfamily of zinc

CC metalloproteases and include aggrecanases amongst their members. The  
 CC protein of the invention demonstrates osteopontin, anti-inflammatory,  
 CC antiarthritic, antirheumatic and cytostatic activities and may be useful  
 CC for treating aggrecanase-associated conditions, including osteoarthritis,  
 CC glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic  
 CC arthritis, periodontal disease and Crohn's disease. The current sequence  
 CC is that of a human aggrecanase modified ADAMTS4 (mTS4) protein of the  
 CC invention.

XX Sequence 646 AA;

Query Match 58.9%; Score 66; DB 8; Length 646;

Best Local Similarity 31.2%; Pred. No. 31;  
 Matches 10; Conservative 12; Mismatches 4; Indels 6; Gaps 1;

Qy 1 WSHPO-----PEKXXXXXXXXXXSHPOFEK 26

Db 615 WLHRRQILIEILRRPFWAGRKGSAMSHPOFEK 646

Search completed: March 2, 2005, 13:02:50

Job time : 118.106 secs

*This Page Blank (uspto)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 80.8537 Seconds  
(without alignments)  
105.489 Million cell updates/sec

Title: SEQ14  
Perfect score: 112  
Sequence: 1 whpghfexkxxxxxxwhpghfexk 26

Scoring table: BLOSUM62GX  
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues  
Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA:\*

1:	/cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep.*
2:	/cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
3:	/cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep.*
4:	/cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep.*
5:	/cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep.*
6:	/cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep.*
7:	/cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep.*
8:	/cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep.*
9:	/cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
10:	/cgn2_6/ptodata/2/pubpaa/US09_PUBCOMB.pep.*
11:	/cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
12:	/cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep.*
13:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
14:	/cgn2_6/ptodata/2/pubpaa/US10_PUBCOMB.pep.*
15:	/cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep.*
16:	/cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
17:	/cgn2_6/ptodata/2/pubpaa/US10D_PUBCOMB.pep.*
18:	/cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
19:	/cgn2_6/ptodata/2/pubpaa/US11_NEW_PUB.pep.*
20:	/cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99	88.4	24	14	US-10-026-578B-3
2	97	86.6	36	14	US-10-026-578B-10
3	84	75.0	36	14	US-10-026-578B-11
4	70	62.5	24	14	US-10-026-578B-4
5	66	58.9	64	16	US-10-628-432-49
6	66	58.9	858	16	US-10-628-432-27
7	64.5	57.6	485	16	US-10-628-432-47
8	64.5	57.6	697	16	US-10-628-432-24
9	64	57.1	24	9	US-09-809-517A-31
10	64	57.1	25	9	US-09-809-517A-34
11	64	57.1	245	17	US-10-887-228A-1
12	64	57.1	246	17	US-10-887-228A-9
13	64	57.1	252	17	US-10-887-228A-5

14	64	57.1	699	16	US-10-408-765A-434	Sequence 434, App
15	64	57.1	699	17	US-10-741-600-851	Sequence 851, App
16	64	57.1	699	17	US-10-741-600-852	Sequence 852, App
17	62	55.4	36	11	US-09-833-245-1103	Sequence 1103, App
18	62	55.4	56	11	US-09-833-245-1105	Sequence 1105, App
19	62	55.4	633	16	US-10-628-432-53	Sequence 53, App
20	61	54.5	21	9	US-09-809-517A-30	Sequence 30, App
21	61	54.5	22	9	US-09-809-517A-33	Sequence 33, App
22	61	54.5	46	15	US-10-432-262-12	Sequence 12, App
23	61	54.5	46	15	US-10-448-609-8	Sequence 8, App
24	61	54.5	117	10	US-09-977-137A-4	Sequence 4, App
25	61	54.5	117	10	US-09-977-137A-5	Sequence 5, App
26	61	54.5	117	10	US-09-977-137A-7	Sequence 7, App
27	61	54.5	117	10	US-09-977-137A-8	Sequence 8, App
28	61	54.5	117	10	US-09-977-137A-9	Sequence 9, App
29	61	54.5	117	10	US-09-977-137A-10	Sequence 10, App
30	61	54.5	117	10	US-09-977-137A-11	Sequence 11, App
31	61	54.5	117	10	US-09-977-137A-12	Sequence 12, App
32	61	54.5	118	10	US-09-977-137A-6	Sequence 6, App
33	61	54.5	661	15	US-10-354-983-8	Sequence 8, App
34	61	54.5	763	15	US-10-358-283-15	Sequence 15, App
35	61	54.5	845	16	US-10-628-432-40	Sequence 40, App
36	60	53.6	513	15	US-10-272-196-39	Sequence 39, App
37	59	52.7	771	15	US-10-369-493-16969	Sequence 16969, A
38	58	51.8	199	16	US-10-437-963-164100	Sequence 164100, A
39	58	51.8	622	15	US-10-424-599-223097	Sequence 223097, A
40	58	51.8	622	15	US-10-425-114-51780	Sequence 51780, A
41	57.5	51.3	134	16	US-10-437-963-104065	Sequence 104065, A
42	57.5	51.3	158	15	US-10-424-599-184769	Sequence 184769, A
43	57.5	51.3	212	15	US-10-424-599-184075	Sequence 184075, A
44	57.5	51.3	252	16	US-10-437-963-200150	Sequence 200150, A
45	57	50.9	121	15	US-10-264-237-2199	Sequence 2199, App

## ALIGNMENTS

RESULT 1  
US-10-026-578B-3  
Sequence 3, Application US/10026578B  
Publication No. US20030083474A1  
GENERAL INFORMATION:  
APPLICANT: IBA (GmbH)  
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta  
FILE REFERENCE: 100810.01US1  
CURRENT APPLICATION NUMBER: US/10/026,578B  
CURRENT FILING DATE: 2002-11-11  
PRIOR APPLICATION NUMBER: DE 101 13 776.1  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: PCT/EP01/11846  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
LENGTH: 24  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Synthetic Peptide  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (3)..(9)  
OTHER INFORMATION: X represents a single amino acid at the position indicated  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (10)..(10)  
OTHER INFORMATION: X represents a single amino acid at the position indicated  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (11)..(11)  
OTHER INFORMATION: X represents a single amino acid at the position indicated  
FEATURE:

```
NAME/KEY: MISC FEATURE
LOCATION: (12)-(12)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (13)-(13)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (14)-(14)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (15)-(15)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC FEATURE
LOCATION: (16)-(16)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: misc feature
OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-3
```

Query Match 86.4%; Score 99; DB 14; Length 24;  
Best Local Similarity 92.3%; Pred. No. 2.4e-05;  
Matches 24; Conservative 0; Mismatches 0; Indels 2; Gaps 1;

Qy 1 WSHPOFEKXXXXXXXXXXWSHPOFEK 26  
Db 1 WSHPOFEK-XXXXXXXXXXWSHPOFEK 24

```
RESULT 2
US-10-026-578B-10
Sequence 10, Application US/10026578B
Publication No. US20030083474A1
GENERAL INFORMATION:
APPLICANT: IBA (GmbH)
APPLICANT: Schmidt, Thomas
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026,578B
PRIOR FILING DATE: 2002-11-11
PRIOR APPLICATION NUMBER: DE 101 13 776.1
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: PCT/EP01/11846
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 10
LENGTH: 36
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
NAME/KEY: MISC FEATURE
LOCATION: (9)-(28)
OTHER INFORMATION: X represents a single amino acid at each of the positions indicat
OTHER INFORMATION: ed; some of the amino acids may be missing. Where amino acids are
FEATURE: missing, the total numbers of x will be no less than 5
NAME/KEY: misc feature
OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-10
```

Query Match 86.6%; Score 97; DB 14; Length 36;  
Best Local Similarity 72.2%; Pred. No. 7.2e-05;  
Matches 26; Conservative 0; Mismatches 0; Indels 10; Gaps 1;

Qy 1 WSHPOFEK-----XXXXXXXXXXWSHPOFEK 26  
Db 1 WSHPOFEK-XXXXXXXXXXWSHPOFEK 24

Db 1 WSHPOFEKXXXXXXXXXXXXXXXXXXXXWSHPOFEK 36

```
RESULT 3
US-10-026-578B-11
Sequence 11, Application US/10026578B
Publication No. US20030083474A1
GENERAL INFORMATION:
APPLICANT: IBA (GmbH)
APPLICANT: Schmidt, Thomas
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026,578B
PRIOR FILING DATE: 2002-11-11
PRIOR APPLICATION NUMBER: DE 101 13 776.1
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: PCT/EP01/11846
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 11
LENGTH: 36
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
NAME/KEY: MISC FEATURE
LOCATION: (9)-(28)
OTHER INFORMATION: Of the five (GGGS) repeats between position 9 and 28, one or more
OTHER INFORMATION: repeats may be missing. However, at least one (GGGS) repeat will
FEATURE: be present
NAME/KEY: misc feature
OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-11
```

Query Match 75.0%; Score 84; DB 14; Length 36;  
Best Local Similarity 41.7%; Pred. No. 0.0035;  
Matches 15; Conservative 10; Mismatches 1; Indels 10; Gaps 1;

Qy 1 WSHPOFEK-----XXXXXXXXXXWSHPOFEK 26  
Db 1 WSHPOFEKGGSGSGSGSGSGSGSGSGSGSGSGSHPOFEK 36

```
RESULT 4
US-10-026-578B-4
Sequence 4, Application US/10026578B
Publication No. US20030083474A1
GENERAL INFORMATION:
APPLICANT: IBA (GmbH)
APPLICANT: Schmidt, Thomas
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026,578B
PRIOR FILING DATE: 2002-11-11
PRIOR APPLICATION NUMBER: DE 101 13 776.1
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: PCT/EP01/11846
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 24
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
NAME/KEY: MISC FEATURE
LOCATION: (9)-(9)
OTHER INFORMATION: X represents a single amino acid at the position indicated
```

```

; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 49
; LENGTH: 646
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: furin-processed construct E
US-10-628-432-49

Query Match          58.9%; Score 66; DB 16; Length 646;
Best Local Similarity 31.2%; Pred. No. 29;
Matches 10; Conservative 12; Mismatches 4; Indels 6; Gaps 1;

Qy      1 WSHPO-----FECKXXXXXXXXXXWSHPQPEK 26
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      615 WLHRRRAQILELIRRRPWAIGKSGAMSHPOPEK 646

RESULT 6
US-10-628-432-27
; Sequence 27, Application US/10628432
; Publication No. US20040142863A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Modified ADAMTS4 molecules
; FILE REFERENCE: AM101378
; CURRENT APPLICATION NUMBER: US/10/628,432
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 27
; LENGTH: 858
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: modified ADAMTS4 molecule
US-10-628-432-27

Query Match          58.9%; Score 66; DB 16; Length 858;
Best Local Similarity 31.2%; Pred. No. 42;
Matches 10; Conservative 12; Mismatches 4; Indels 6; Gaps 1;

Qy      1 WSHPO-----FECKXXXXXXXXXXWSHPQPEK 26
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      827 WLHRRRAQILELIRRRPWAIGKSGAMSHPOPEK 858

RESULT 7
US-10-628-432-47
; Sequence 47, Application US/10628432
; Publication No. US20040142863A1
; GENERAL INFORMATION:
; APPLICANT: Wyeth
; TITLE OF INVENTION: Modified ADAMTS4 molecules
; FILE REFERENCE: AM101378
; CURRENT APPLICATION NUMBER: US/10/628,432
; CURRENT FILING DATE: 2003-07-29
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 47
; LENGTH: 485
; TYPE: PRT
; ORGANISM: Artificial
; FEATURE:
; OTHER INFORMATION: furin-processed construct C
US-10-628-432-47

Query Match          57.6%; Score 64.5; DB 16; Length 485;
Best Local Similarity 36.7%; Pred. No. 32;
Matches 11; Conservative 12; Mismatches 2; Indels 5; Gaps 1;

Qy      2 SHPOPEK-----XXXXXXXXXXWSHPQPEK 26
      | | | | | | | | | | | | | | | | | | | | | | | | | | | |
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 10
US-09-809-517A-34
; Sequence 34, Application US/09809517A
; Patent No. US20020034733A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: NO. US20020034733A1
methode for displaying (poly)peptides/prote

```

```

1  APPLICANT: Schering AG
2  TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio
3  TITLE OR INVENTION: Substance That is Contained in a Sample
4  FILE REFERENCE: 530569US
5  CURRENT APPLICATION NUMBER: US/10/867,228A
6  CURRENT FILING DATE: 2004-07-09
7  PRIOR APPLICATION NUMBER: DE 103 31 093.2

```

```
/ PRIOR FILING DATE: 2003-07-09
/ PRIOR APPLICATION NUMBER: US 60/478,262
/ PRIOR FILING DATE: 2003-07-16
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: Patent version 3.1
/ SEQ ID NO 9
/ LENGTH: 246
/ TYPE: PRT
/ ORGANISM: artificial
/ FEATURE:
/ OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-9

Query Match          57.1%; Score 64; DB 17; Length 246;
Best Local Similarity 41.7%; Pred. No. 16;
Matches 10; Conservative 12; Mismatches 0; Indels 2; Gaps 1;

Qy 5 QFEK-XXXXXXXXXXWSHPQFEK 26
Db 223 EFEQKLISEEDLNGAPWMSHPQFEK 246

RESULT 13
US-10-887-228A-5
/ Sequence 5, Application US/10887228A
/ Publication No. US20050037402A1
/ GENERAL INFORMATION:
/ APPLICANT: Schering AG
/ TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio
/ FILE REFERENCE: S30569US
/ CURRENT APPLICATION NUMBER: US/10/887,228A
/ PRIOR FILING DATE: 2004-07-09
/ PRIOR APPLICATION NUMBER: DE 103 31 093.2
/ PRIOR FILING DATE: 2003-07-09
/ PRIOR APPLICATION NUMBER: US 60/478,262
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: Patent version 3.1
/ SEQ ID NO 5
/ LENGTH: 252
/ TYPE: PRT
/ ORGANISM: artificial
/ FEATURE:
/ OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-5

Query Match          57.1%; Score 64; DB 17; Length 252;
Best Local Similarity 41.7%; Pred. No. 16;
Matches 10; Conservative 12; Mismatches 0; Indels 2; Gaps 1;

Qy 5 QFEK-XXXXXXXXXXWSHPQFEK 26
Db 229 EFEQKLISEEDLNGAPWMSHPQFEK 252

RESULT 14
US-10-408-765A-434
/ Sequence 434, Application US/10408765A
/ Publication No. US20040101874A1
/ GENERAL INFORMATION:
/ APPLICANT: Ghosh, Soumitra S.
/ APPLICANT: Pany, Bin D.
/ APPLICANT: Zhang, Bing
/ APPLICANT: Gibson, Bradford W.
/ APPLICANT: Taylor, Steven W.
/ APPLICANT: Glenn, Gary M.
/ APPLICANT: Warnock, Dale E.
/ TITLE OF INVENTION: TARGETS FOR THERAPEUTIC INTERVENTION
/ FILE REFERENCE: 660088.465
/ CURRENT APPLICATION NUMBER: US/10/408,765A
/ CURRENT FILING DATE: 2003-04-04
```

```
/ NUMBER OF SEQ ID NOS: 3077
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 434
/ LENGTH: 699
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-408-765A-434

Query Match          57.1%; Score 64; DB 16; Length 699;
Best Local Similarity 37.5%; Pred. No. 59;
Matches 9; Conservative 10; Mismatches 5; Indels 0; Gaps 0;

Qy 3 HPQEKXXXXXXXXXXWSHPQFEK 26
Db 67 HPDENKDVSTENGHHFWSHPDREK 90

RESULT 15
US-10-741-600-851
/ Sequence 851, Application US/10741600
/ Publication No. US20050026169A1
/ GENERAL INFORMATION:
/ APPLICANT: CARGILL, Michele et al.
/ TITLE OF INVENTION: GENETIC POLYMORPHISMS ASSOCIATED WITH
/ FILE REFERENCE: CI001499
/ CURRENT APPLICATION NUMBER: US/10/741,600
/ PRIOR FILING DATE: 2003-12-22
/ NUMBER OF SEQ ID NOS: 73997
/ SOFTWARE: FastSeq for Windows Version 4.0.
/ SEQ ID NO 851
/ LENGTH: 699
/ TYPE: PRT
/ ORGANISM: Homo sapiens
US-10-741-600-851

Query Match          57.1%; Score 64; DB 17; Length 699;
Best Local Similarity 37.5%; Pred. No. 59;
Matches 9; Conservative 10; Mismatches 5; Indels 0; Gaps 0;

Qy 3 HPQEKXXXXXXXXXXWSHPQFEK 26
Db 67 HPDENKDVSTENGHHFWSHPDREK 90

Search completed: March 2, 2005, 14:19:02
Job time : 80.8537 secs
```

*This Page Blank (uspto)*



```
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-517A-34

Query Match          57.1%; Score 64; DB 4; Length 25;
Best Local Similarity 41.7%; Pred. NO. 0.077;
Matches 10; Conservative 12; Mismatches 0; Indels 2; Gaps 1;

QY 5 QFEK-XXXXXXXXXXWSHPQFEK 26
   |||:|||||:|||||:|||||
DB 2 EFDKLISEDDNGAPWWSHPQFEK 25

RESULT 3
US-09-538-092-995
; Sequence 995, Application US/09538092
; Patent No. 6753314
; GENERAL INFORMATION:
; APPLICANT: Glot, Loic
; APPLICANT: Mansfield, Traci A.
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same
; FILE REFERENCE: 15966-542
; CURRENT APPLICATION NUMBER: US/09/538,092
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: 60/127,352
; PRIOR FILING DATE: 1999-04-01
; PRIOR APPLICATION NUMBER: 60/178,965
; PRIOR FILING DATE: 2000-02-01
; NUMBER OF SEQ ID NOS: 1387
; SOFTWARE: CuratSeqFormatter Version 0.9
; SEQ ID NO 995
; LENGTH: 699
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (0)..(0)
; OTHER INFORMATION: Polypeptide Accession Number P23327
US-09-538-092-995

Query Match          57.1%; Score 64; DB 4; Length 699;
Best Local Similarity 37.5%; Pred. NO. 9.1;
Matches 9; Conservative 10; Mismatches 5; Indels 0; Gaps 0;

QY 3 HPOFEKXXXXXXXXXXWSHPQFEK 26
   |||:|||||:|||||:|||||
DB 67 HPDENKDVSTENGHHFWWSHPDREK 90

RESULT 4
US-09-809-517A-30.
; Sequence 30, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on h
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 30
; LENGTH: 21
```

```
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-517A-30

Query Match          54.5%; Score 61; DB 4; Length 21;
Best Local Similarity 44.4%; Pred. NO. 0.16;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

QY 9 XXXXXXXXXXXXXWSHPQFEK 26
   |||:|||||:|||||:|||||
DB 4 YKDDDKGAPWWSHPQFEK 21

RESULT 5
US-09-809-517A-33
; Sequence 33, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 33
; LENGTH: 22
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33

Query Match          54.5%; Score 61; DB 4; Length 22;
Best Local Similarity 44.4%; Pred. NO. 0.17;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

QY 9 XXXXXXXXXXXXXWSHPQFEK 26
   |||:|||||:|||||:|||||
DB 5 YKDDDKGAPWWSHPQFEK 22

RESULT 6
US-09-977-137A-4
; Sequence 4, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Casagiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4
```



Query Match Similarity	54.5%	Score 61	DB 4	Length 117	Best Local Similarity	44.4%	Pred. No. 1.9	Matches	8	Conservative	10	Mismatches	0	Indels	0	Gaps	0
Query	9 XXXXXXXXXXXXSHPOFEK 26																
DB	100 RKGNVSCPSAWSHPOFEK 117																
<p>RESULT 7</p> <p>US-09-977-137A-5</p> <p>Sequence 5, Application US/09977137A</p> <p>Patent No. 6750042</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Sumner, Anne O.</p> <p>APPLICANT: Caguiat, Jonathan</p> <p>TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and</p> <p>TITLE OF INVENTION: Methods</p> <p>FILE REFERENCE: 79-00</p> <p>CURRENT APPLICATION NUMBER: US/09/977,137A</p> <p>CURRENT FILING DATE: 2001-10-12</p> <p>PRIOR APPLICATION NUMBER: US 60/240,465</p> <p>PRIOR FILING DATE: 2000-10-12</p> <p>NUMBER OF SEQ ID NOS: 18</p> <p>SOFTWARE: PatentIn Ver. 2.0</p> <p>SEQ ID NO 5</p> <p>LENGTH: 117</p> <p>TYPE: PRT</p> <p>ORGANISM: Artificial Sequence</p> <p>FEATURE:</p> <p>OTHER INFORMATION: Description of Artificial Sequence: chelon</p> <p>US-09-977-137A-5</p>																	
Query Match	54.5%	Score 61	DB 4	Length 117	Best Local Similarity	44.4%	Pred. No. 1.9	Matches	8	Conservative	10	Mismatches	0	Indels	0	Gaps	0
DB	100 RKGNVSCPSAWSHPOFEK 117																
<p>RESULT 8</p> <p>US-09-977-137A-7</p> <p>Sequence 7, Application US/09977137A</p> <p>Patent No. 6750042</p> <p>GENERAL INFORMATION:</p> <p>APPLICANT: Sumner, Anne O.</p> <p>APPLICANT: Caguiat, Jonathan</p> <p>TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and</p> <p>TITLE OF INVENTION: Methods</p> <p>FILE REFERENCE: 79-00</p> <p>CURRENT APPLICATION NUMBER: US/09/977,137A</p> <p>CURRENT FILING DATE: 2001-10-12</p> <p>PRIOR APPLICATION NUMBER: US 60/240,465</p> <p>PRIOR FILING DATE: 2000-10-12</p> <p>NUMBER OF SEQ ID NOS: 18</p> <p>SOFTWARE: PatentIn Ver. 2.0</p> <p>SEQ ID NO 7</p> <p>LENGTH: 117</p> <p>TYPE: PRT</p> <p>ORGANISM: Artificial Sequence</p> <p>FEATURE:</p> <p>OTHER INFORMATION: Description of Artificial Sequence: chelon</p> <p>US-09-977-137A-7</p>																	
Query Match	54.5%	Score 61	DB 4	Length 117	Best Local Similarity	44.4%	Pred. No. 1.9	Matches	8	Conservative	10	Mismatches	0	Indels	0	Gaps	0
DB	100 RKGNVSCPSAWSHPOFEK 117																
<p>Query Match</p> <p>Best Local Similarity</p> <p>Matches</p> <p>8</p> <p>Conservative</p> <p>10</p> <p>Mismatches</p> <p>0</p> <p>Indels</p> <p>0</p> <p>Gaps</p> <p>0</p>																	

```

RESULT 9
US-09-977-137A-8
; Sequence 8, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 8
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-8

Query Match          54.5%; Score 61; DB 4; Length 117;
Best Local Similarity 44.4%; Pred. No. 1.9;
Matches      8; Conservative 10; Mismatches      0; Indels      0; Gaps      0;

QY      9 XXXXXXXXXXXXSHPOFEK 26
          :XXXXXXXXXXXXXXXXXXXX
DB      100 RKNVSCPSANSHPOFEK 117

RESULT 10
US-09-977-137A-9
; Sequence 9, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-9

Query Match          54.5%; Score 61; DB 4; Length 117;
Best Local Similarity 44.4%; Pred. No. 1.9;
Matches      8; Conservative 10; Mismatches      0; Indels      0; Gaps      0;

QY      9 XXXXXXXXXXXXSHPOFEK 26
          :XXXXXXXXXXXXXXXXXXXX
DB      100 RKNVSCPSANSHPOFEK 117

RESULT 11
US-09-977-137A-10
; Sequence 10, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:

```

```

; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-10
```

```

Query Match          54.5%; Score 61; DB 4; Length 117;
Best Local Similarity 44.4%; Pred. No. 1.9;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 9 XXXXXXXXXXXXSHPOFEK 26
   :XXXXXXXXXXXXXXXXXXXX
DB 100 RKGWVCPGSAWSHPOFEK 117
```

```

RESULT 12
US-09-977-137A-11
; Sequence 11, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-11
```

```

Query Match          54.5%; Score 61; DB 4; Length 117;
Best Local Similarity 44.4%; Pred. No. 1.9;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 9 XXXXXXXXXXXXSHPOFEK 26
   :XXXXXXXXXXXXXXXXXXXX
DB 100 RKGWVCPGSAWSHPOFEK 117
```

```

RESULT 13
US-09-977-137A-12
; Sequence 12, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
```

```

; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 12
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-12
```

```

Query Match          54.5%; Score 61; DB 4; Length 117;
Best Local Similarity 44.4%; Pred. No. 1.9;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 9 XXXXXXXXXXXXSHPOFEK 26
   :XXXXXXXXXXXXXXXXXXXX
DB 100 RKGWVCPGSAWSHPOFEK 117
```

```

RESULT 14
US-09-977-137A-6
; Sequence 6, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 6
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-6
```

```

Query Match          54.5%; Score 61; DB 4; Length 118;
Best Local Similarity 44.4%; Pred. No. 1.9;
Matches 8; Conservative 10; Mismatches 0; Indels 0; Gaps 0;
```

```

QY 9 XXXXXXXXXXXXSHPOFEK 26
   :XXXXXXXXXXXXXXXXXXXX
DB 101 RKGWVCPGSAWSHPOFEK 118
```

```

RESULT 15
US-08-770-761A-2
; Sequence 2, Application US/08770761A
; Patent No. 5814503
; GENERAL INFORMATION:
; APPLICANT: Kovacevic, Steven
; APPLICANT: Otto, Keith A.
; APPLICANT: Rao, Ramachandra N.
; TITLE OF INVENTION: FUSION PROTEINS COMPRISING CELL CYCLE
; TITLE OF INVENTION: REGULATORY PROTEINS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center/Patent Division
; CITY: Indianapolis
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
```



*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 20.6341 Seconds  
(without alignments)  
125.901 Million cell updates/sec

Title: SEQ15  
Perfect score: 113  
Sequence: 1 wshpqfexkxxxxxxxxxwshpqfex 27

Scoring table: BLOSUM62GX  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
1: p1r1:\*  
2: p1r2:\*  
3: p1r3:\*  
4: p1r4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	61	54.0	433	2 H83444	probable cytochrome
2	59	52.2	219	2 C96510	hypothetical prote
3	58	51.3	103	2 E90973	hypothetical prote
4	58	51.3	103	2 A85821	unknown protein en
5	58	51.3	103	2 C85713	unknown protein en
6	58	51.3	103	2 A90904	hypothetical prote
7	57.5	50.9	699	2 A54660	histidine rich cal
8	56	49.6	242	2 T40523	hypothetical prote
9	55	48.7	207	2 JH0145	hypothetical 24.1k
10	55	48.7	224	2 S71749	DCI. protein precu
11	54.5	48.2	1087	2 T22847	hypothetical prote
12	54.5	48.2	1250	2 T22845	hypothetical prote
13	54	47.8	211	2 B71058	hypothetical prote
14	54	47.8	540	2 T10892	probable calnexin
15	54	47.8	600	2 I49281	ferillin alpha pre
16	54	47.8	947	2 A12787	valyl-tRNA synthet
17	54	47.8	947	2 B97567	valyl-tRNA synthet
18	54	47.8	2037	2 S37178	fatty-acyl-CoA syn
19	53.5	47.3	304	2 A81069	mr restriction sy
20	53.5	47.3	304	2 A40368	mr protein - Bsch
21	53	46.9	367	2 F64202	probable GTP-Dindi
22	53	46.9	403	2 A12363	transposase all446
23	53	46.9	403	2 AE2518	transposase all732
24	53	46.9	421	2 G82422	anaerobic glycerol
25	53	46.9	421	2 A47713	chitin deacetylase
26	53	46.9	453	2 D95342	Nosd periplasmic c
27	53	46.9	478	2 T40362	concerned hypochet
28	53	46.9	684	2 S52835	hypothetical prote
29	53	46.9	812	1 A36477	fibroblast growth

30	53	46.9	814	1 A39752	fibroblast growth
31	53	46.9	909	2 A03419	phage host specific
32	53	46.9	1209	2 T31657	reverse transcript
33	53	46.9	2471	2 T42977	large tegument pro
34	52.5	46.5	342	2 S37596	protein kinase MCP
35	52	46.0	332	2 I39927	ABC transporter (b
36	52	46.0	334	2 AF1151	6-phosphogluconate
37	52	46.0	334	2 AF1151	probable 6-phospho
38	52	46.0	362	2 C90674	hypothetical prote
39	52	46.0	362	2 C64758	Yahp protein - Bsc
40	52	46.0	365	2 C85524	hypothetical prote
41	52	46.0	391	2 G86718	replication protei
42	52	46.0	403	2 AB1951	transposase alr115
43	52	46.0	449	2 S07714	T64 protein precu
44	52	46.0	450	2 E70590	3-phosphoshikimate
45	52	46.0	451	2 I50131	clusterin - quail

## ALIGNMENTS

```

RESULT 1
H83444
probable cytochrome c P1660 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 09-Jul-2004
C:Accession: H83444
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kae, A.; Larbig, K.; Lam,
., J.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A>Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic path
A:Reference number: A02950; PMID:20437337; PMID:10984043
A:Accession: H83444
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-433 <STO-
A:Cross-references: UNIPROT:Q913C1; GB:AE004588; GB:AE004091; NID:99947563; PIDN:AMG0491
A:Experimental source: strain PA01
C:Genetics:
A:Gene: P1660
C:Superfamily: membrane-bound alcohol dehydrogenase, cytochrome c subunit; cytochrome c
C:Keywords: chromoprotein; heme; iron; metalloprotein
F:58,61/Binding site: heme (Cys) (covalent) #status predicted
F:62/Binding site: heme iron (His) (axial ligand) #status predicted
F:204,207/Binding site: heme (Cys) (covalent) #status predicted
F:208/Binding site: heme iron (His) (axial ligand) #status predicted
F:336,339/Binding site: heme (Cys) (covalent) #status predicted
F:340/Binding site: heme iron (His) (axial ligand) #status predicted

Query Match          54.0%; Score 61; DB 2; Length 433;
Best Local Similarity 28.0%; Pred. No. 8.7;
Matches 7; Conservative 15; Mismatches 3; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXWSHPQF 25
DB 99 WSPAFERMRHGVARDGSYLPAP 123

RESULT 2
hypothetical protein P2919.1 [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C>Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 09-Jul-2004
C:Accession: C96510
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federjfel, N.A.; Kaul, S.; White, O.; Alonso,
Chin, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.;
ansen, N.F.; Hughes, B.; Huizar, L.
Nature 408, 816-820, 2000
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.
C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Malt, R.; Marziani,
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon,

```

```

ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
A>Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86144; MUID:21016719; PMID:11130712
A:Accession: C96510
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-219 <STO>
C:Cross-references: UNIPROT:Q9C642; GB:AE005173; NID:g11321777; PIDN:AAG34254.1; GSPDB:C
C:Genetics:
A:Gene: P3G19.1
A:Map position: 1
C:Superfamily: tomato DCL protein

Query Match          52.2%; Score 59; DB 2; Length 219;
Best Local Similarity 33.3%; Pred. No. 7.5;
Matches      8; Conservative 12; Mismatches    4; Indels    0; Gaps    0;

Db
   3 HPOFEKXXXXXXXXXXWSHPQFE 26
     |||:::|||||||:::|||||
  146 HPECEKKIGCGIDYIMVGHHPDFE 169

RESULT 3
E90973
Hypothetical protein EC62757 [imported] - Escherichia coli (strain O157:H7, substrain R)
C:Species: Escherichia coli
C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:Accession: E90973
R:Hayashi, T.; Makino, K.; Onishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, I.; Hattori, W.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A>Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: E90973
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <HAV>
C:Cross-references: UNIPROT:Q8X4V0; GB:BA000007; PIDN:EAB36180.1; PID:g13362225; GSPDB:C
C:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: EC62757

Query Match          51.3%; Score 58; DB 2; Length 103;
Best Local Similarity 30.0%; Pred. No. 4.3;
Matches      6; Conservative 14; Mismatches    0; Indels    0; Gaps    0;

Cy
   8 KXXXXXXXWWSHPQFEK 27
     |:::|||||:::|||||
  5 KEMPEVERNEYGCWTHPEYEK 24

RESULT 4
A85821
Unknown protein encoded within prophage CP-933U [imported] - Escherichia coli (strain O157:H7)
C:Species: Escherichia coli
C>Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:Accession: A85821
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew
Miller, L.; Grobeck, E.O.; Davis, N.W.; Lim, A.; Diallanta, E.; Potamousis, K.; Apodaca,
Native 409, 529-533, 2001
A>Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:Reference number: A85480; MUID:21074935; PMID:11206551
A:Accession: A85821
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-103 <STO>
C:Cross-references: UNIPROT:Q8X4V0; GB:AE005174; NID:g12516136; PIDN:AAG57029.1; GSPDB:C
C:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: Z3120

```

```

Matches      6; Conservative    14; Mismatches     0; Indels       0; Gaps        0;

QY          8 KXXXXXXXXXXMSHPQFEK 27
|:::|||||||::|
DB          5 KEMPVVRDGYGWTHTPEYEK 24

RESULT 5
Unknown protein encoded within prophage CP-9330 [imported] - Escherichia coli (strain O1
C85713
C:/Species: Escherichia coli
C:/Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 09-Jul-2004
C:/Accession: C85713
R:/Perna, N.T., Plunkett III, G.; Burland, V.; Mau, B.; Glaeser, J.D.; Rose, D.J.; Mayhew
iller, L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamouzis, K.; Apodaca,
Nature 409, 529-533, 2001
A:/Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.
A:/Reference number: A85480; MUID:21074935; PMID:11206551
A:/Accession: C85713
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-103 <STO>
A:/Cross-references: UNIPROT:Q8X549; GB:AE005174; NID:g12515046; PIDN:AAG56167.1; GSPDB:G
A:/Experimental source: strain O157:H7, substrain EDL933
C:/Genetics:
A:/Gene: Z2097

Query Match      51.3%; Score 58; DB 2; Length 103;
Best Local Similarity 30.0%; Pred. No. 4.3;
Matches      6; Conservative    14; Mismatches     0; Indels       0; Gaps        0;

QY          8 KXXXXXXXXXXMSHPQFEK 27
|:::|||||||::|
DB          5 KEMPVVRDGYGWTHTPEYEK 24

RESULT 6
Hypothetical protein ECs2201 [imported] - Escherichia coli (strain O157:H7, substrain R1.
A90904
C:/Species: Escherichia coli
C:/Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 09-Jul-2004
C:/Accession: A90904
R:/Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
DNA Res. 8, 11-22, 2001
A:/Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and geno-
A:/Reference number: A99629; MUID:21156231; PMID:11258796
A:/Accession: A90904
A:/Status: preliminary
A:/Molecule type: DNA
A:/Residues: 1-103 <HAV>
A:/Cross-references: UNIPROT:Q8X549; GB:BA000007; PIDN:BA835624.1; PID:g13361667; GSPDB:G
A:/Experimental source: strain O157:H7, substrain RIMD 0509952
C:/Genetics:
A:/Gene: ECs2201

Query Match      51.3%; Score 58; DB 2; Length 103;
Best Local Similarity 30.0%; Pred. No. 4.3;
Matches      6; Conservative    14; Mismatches     0; Indels       0; Gaps        0;

QY          8 KXXXXXXXXXXMSHPQFEK 27
|:::|||||||::|
DB          5 KEMPVVRDGYGWTHTPEYEK 24

RESULT 7
histidine rich calcium binding protein - human
A54660
C:/Species: Homo sapiens (man)
C:/Date: 02-Jun-1995 #sequence_revision 02-Jun-1995 #text_change 09-Jul-2004
C:/Accession: A54660
R:/Hotmann, S.L.; Topham, M.; Hsieh, C.L.; Francke, U.
denomies 9, 656-669, 1991

```

```

A>Title: cDNA and genomic cloning of HRC, a human sarcoplasmic reticulum protein, and 1c
A|Reference number: A54660; MUID:91244309; PMID:2037293
A|Accession: A54660
A|Status: preliminary
A|Molecule type: mRNA
A|Residues: 1-699 <HOP>
A|Cross-references: UNIPROT:P23327; GB:M60052; NID:g183918; PIDN:AAA88071.1; PID:g183919
C|GeneID:b
A|Gene: GDB:HRC
A|Cross-references: GDB:I26369; OMIM:142705
A|Map position: 19q13.3-19q13.3
C|Keywords: calcium binding

Query Match      50.9%; Score 57.5; DB 2; Length 699;
Best Local Similarity 36.0%; Pred. No. 47;
Matches 9; Conservative 11; Mismatches 4; Indels 1; Gaps 1;

Oy      3 HPQFEKXXXXXXXXXXWSHPQFEK 27
        ||| | : : : : : : : |||
Db      67 HPD-ENKDVSTENGHHFMSHPREX 90

RESULT 8
T40523
hypothetical protein SPBC530.07c - fission yeast (Schizosaccharomyces pombe)
C|Species: Schizosaccharomyces pombe
C|Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 09-Jul-2004
C|Accession: T40523
R|Lyne, M.; Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.
submitted to the EMBL Data Library, May 1998
A|Reference number: Z21934
A|Accession: T40523
A|Status: preliminary; translated from GB/EMBL/DDBJ
A|Molecule type: DNA
A|Residues: 1-242 <LYN>
A|Cross-references: UNIPROT:O59743; EMBL:AL023634; PIDN:CAA19173.1; GSPDB:GN00067; SPDB:
A|Experimental source: strain 972h-, cosmid c530
C|GeneID:c
A|Gene: SPDB:SPBC530.07C
A|Map position: 2

Query Match      49.6%; Score 56; DB 2; Length 242;
Best Local Similarity 30.4%; Pred. No. 22;
Matches 7; Conservative 12; Mismatches 4; Indels 0; Gaps 0;

Oy      5 QPEKXXXXXXXXXXWSHPQFEK 27
        ||| | : : : : : : : |||
Db     180 QPKNCPSEMEVKKWGPSPFK 202

RESULT 9
JH0145
hypothetical 24.1k protein - Neurospora crassa
C|Species: Neurospora crassa
C|Date: 31-Dec-1991 #sequence_revision 31-Dec-1991 #text_change 09-Jul-2004
C|Accession: JH0145
R|Schechtman, M.G.
Gene 88, 159-165, 1990
A|Title: Characterization of telomere DNA from Neurospora crassa.
A|Reference number: JH0145; MUID:90269603; PMID:1971801
A|Accession: JH0145
A|Molecule type: DNA
A|Residues: 1-207 <SCH>
A|Cross-references: UNIPROT:O01400; GB:M37064; NID:g168912; PIDN:AAC13656.1; PID:g529466
A|Experimental source: strain OR23-IV-A
A|Note: translation of the nucleotide sequence is not complete
C|Superfamily: Neurospora crassa hypothetical 24.1k protein

Query Match      48.7%; Score 55; DB 2; Length 207;
Best Local Similarity 30.0%; Pred. No. 26;
Matches 6; Conservative 13; Mismatches 1; Indels 0; Gaps 0;

Oy      6 FEKXXXXXXXXXXWSHPQF 25

```

```

Db          119 FEALITSHQYRSLMNHPRF 138
|||||.....|||.|||
RESULT 10
S71749
DCL protein precursor, chloroplast - tomato
C:Species: Lycopersicon esculentum (tomato)
C:Date: 29-Jan-1998 #sequence_revision 13-Feb-1998 #text_change 09-Jul-2004
C:Accession: S71749; S71748
R:Keddie, J.S.; Carroll, B.; Jones, J.D.G.; Griseem, W.
EMBO J. 15, 4208-4217, 1996
A>Title: The DCL gene of tomato is required for chloroplast development and palisade cell
A:Reference number: S71748; MUID:97015121; PMID:8861949
A:Accession: S71749
A>Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-224 <KED>
A:Cross-references: UNIPROT:Q42463; EMBL:U55278; NID:G1323697; PID:G1323698
A:Experimental source: strain Moneymaker
A>Note: The nucleotide sequence was submitted to the EMBL Data Library, April 1996
A:Accession: S71748
A:Molecule type: mRNA
A:Residues: 1-224 <KEN>
A:Cross-references: EMBL:U55219; NID:G1305530; PIDD:AAC49433.1; PID:G1305531
A:Experimental source: strain Moneymaker
A:Genetics:
A:Map position: 4
A:Genome: nuclear
A:Introns: 128/2; 166/3
C:Function:
A>Description: required for both chloroplast development and palisade cell morphogenesis
C:Superfamily: tomato DCL protein
C:Keywords: chloroplast
F:1-50/Domain: transit peptide (chloroplast) #status predicted <TMP>
F:51-224/Product: DCL protein #status predicted <MAT>

Cy      3 HQPERXXXXXXXXXXMSHPOFE 26
||:::||:::||:::||
Query Match           48.7%; Score 55; DB 2; Length 224;
Best Local Similarity 29.2%; Pred. No. 28;
Matches 7; Conservative 13; Mismatches 4; Indels 0; Gaps 0;

Db      151 HPECCKIIGPVDYITVGYPDFE 174

RESULT 11
T22847
hypochlorite protein F57C7.lb - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 09-Jul-2004
C:Accession: T22847
R:White, S.
Submitted to the EMBL Data Library, February 1996
A:Reference number: Z19625
A:Accession: T22847
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1087 <WIL>
A:Cross-references: UNIPROT:Q20948; EMBL:D69646; PIDD:CAA93475.1; GSPDB:GN00028; CESP:F5
A:Experimental source: clone F57C7
A:Genetics:
A:Gene: CESP:F57C7.lb
A:Map position: X
A:Introns: 262/3; 351/3; 391/3; 627/3; 795/2; 957/1
F:307-364/Domain: bromodomain homology <BRO1>
F:579-636/Domain: bromodomain homology <BRO2>

Query Match           48.2%; Score 54.5; DB 2; Length 1087;
Best Local Similarity 22.9%; Pred. No. 21e+02;
Matches 8; Conservative 15; Mismatches 3; Indels 9; Gaps 1;

2 SHPOFEKXXXXXXXXXXXXWSH-----POFEK 27

```

Db 524 AHQETPRKKNPTLIEWKHLVPRWQGIPEWQK 558

## RESULT 12

T22845

hypothetical protein F57C7.1a - *Caenorhabditis elegans*

C/Species: *Caenorhabditis elegans*

C/Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C/Accession: T22845

R/White, S.

Submitted to the EMBL Data Library, February 1996

A/Reference number: Z19625

A/Accession: T22845

A/Status: preliminary; translated from GB/EMBL/DBD

A/Molecule type: DNA

A/Residues: 1-1250 <MTL>

A/Cross-references: UNIPROT:Q020947; EMBL:Z69646; PIDN:CAA93473.1; GSPDB:GN00026; CESP:FS

A/Experimental source: clone F57C7

C/Genetics:

A/Map position: X

A/Genes: CESP:F57C7.1a

A/Intons: 262/3; 351/3; 391/3; 431/3; 667/3; 835/2; 997/1; 1127/1; 1218/3

F/307-364/Domain: bromodomain homology <BRO1>

F/619-676/Domain: bromodomain homology <BRO2>

## Query Match

Best Local Similarity 48.2%; Score 54.5; DB 2; Length 1250;

Matches 8; Conservative 15; Mismatches 3; Indels 9; Gaps 1;

QY 2 SHPQFEKXXXXXXXXXWSH-----PQFEK 27

Db 564 AHQETPRKKNPTLIEWKHLVPRWQGIPEWQK 598

## RESULT 13

B71058

hypothetical protein PH160 - *Pyrococcus horikoshii*

C/Species: *Pyrococcus horikoshii*

C/Date: 14-Aug-1998 #sequence\_revision 14-Aug-1998 #text\_change 12-Jul-2004

C/Accession: B71058

R/Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Halkawa, Y.; Hino, Y.; Yamamoto, S.; Sekin

M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kuehida, N.; Oguchi

DNA Res. 5, 55-76, 1998

A/Title: Complete sequence and gene organization of the genome of a hyper-thermophilic e

A/Reference number: A71000; MUID:98344337; PMID:9679194

A/Accession: B71058

A/Status: preliminary; nucleic acid sequence not shown; translation not shown

A/Molecule type: DNA

A/Residues: 1-211 <KAW>

A/Cross-references: UNIPROT:Q58874; GB:AP000005; NID:G3236132; PIDN:BAA30260.1; PID:G325

A/Experimental source: strain OT3

A/Note: this accession replaces an interim accession for a sequence replaced by GenBank

C/Genetics:

A/Genes: PH160

## Query Match

Best Local Similarity 47.8%; Score 54; DB 2; Length 211;

Matches 9; Conservative 14; Mismatches 4; Indels 2; Gaps 1;

QY 1 WSHPQFEKXXXXXXXXXWSH--HPQFEK 27

Db 158 WSHPQFEKXXXXXXXXXWSH--HPQFEK 186

## RESULT 14

T10892

probable calnexin - *Jerusalem artichoke*

C/Species: *Helianthus tuberosus* (*Jerusalem artichoke*)

C/Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004

C/Accession: T10892

R/Haefenratz, M.; Jeltsch, J.; Lesot, A.; Michalak, M.; Durst, F.

Submitted to the EMBL Data Library, July 1994

A/Description: Cloning and characterization of a cDNA encoding an analog of the calnexin

A/Reference number: Z17201

A/Accession: T10892

A/Status: preliminary; translated from GB/EMBL/DBD

A/Molecule type: mRNA

A/Residues: 1-540 <HAS>

A/Cross-references: UNIPROT:Q39994; EMBL:Z35108; NID:G510906; PID:G510907

A/Experimental source: cv. blanc commun; tuber slice; parenchyma

C/Function:

A/Description: ER-bound chaperone

C/Superfamily: calnexin

C/Keywords: endoplasmic reticulum; molecular chaperone

## Query Match

Best Local Similarity 47.8%; Score 54; DB 2; Length 540;

Matches 6; Conservative 11; Mismatches 6; Indels 0; Gaps 0;

QY 1 WSHPQFEKXXXXXXXXXWSHP 23

Db 301 WSHPQFEKXXXXXXXXXWSHP 323

## RESULT 15

I49281

fertilin alpha precursor - mouse (fragment)

C/Species: *Mus musculus* (house mouse)

C/Date: 02-Jul-1996 #sequence\_revision 02-Jul-1996 #text\_change 16-Aug-2004

C/Accession: I49281

R/Wolfsberg, T.G.; Stralight, P.D.; Gerena, R.L.; Huovila, A.P.; Primakoff, P.; Wyle, D.

Dev. Biol. 169, 378-383, 1995

A/Title: ADAM, a widely distributed and developmentally regulated gene family encoding m

A/Reference number: I48100; MUID:95268891; PMID:7750654

A/Accession: I49281

A/Status: preliminary; translated from GB/EMBL/DBD

A/Molecule type: mRNA

A/Residues: 1-600 <RES>

A/Cross-references: EMBL:U22056; NID:G965009; PIDN:AAA74920.1; PID:G965010

C/Genetics:

A/Genes: ADAM 1

C/Superfamily: disintegrin homology

F/246-326/Domain: disintegrin homology <DIS>

F/180/Active site: Glu #status predicted

## Query Match

Best Local Similarity 47.8%; Score 54; DB 2; Length 600;

Matches 6; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

QY 2 SHPQFEKXXXXXXXXXWSHPQF 25

Db 22 SHPQFEKXXXXXXXXXWSHPQF 45

Search completed: March 2, 2005, 12:29:01  
Job time : 21.6341 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 100.317 Seconds  
(without alignment)  
137.824 Million cell updates/sec

Title: SEQ15  
Perfect score: 113  
Sequence: 1 whpbfekxxxxxxxxxxxxwhpbfek 27

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues

Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 03: \*  
1: uniprot\_sprot: \*  
2: uniprot\_trembl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	68	60.2	107	2	062315
2	67	59.3	704	2	08CJD2
3	62	54.9	743	2	08CH90
4	61	54.0	433	2	0913C1
5	60	53.1	224	2	08RVH8
6	59	52.2	219	2	09C642
7	59	52.2	389	2	09AD61
8	58.5	51.8	208	2	06ZFR3
9	58.5	51.8	256	2	0651B2
10	58	51.3	103	2	08X549
11	58	51.3	103	2	08X4V0
12	58	51.3	366	2	08BMQ4
13	58	51.3	398	2	06V551
14	58	51.3	448	2	08XND2
15	57.5	50.9	699	1	SRCH_HUMAN
16	57	50.4	158	2	072942
17	57	50.4	294	2	07US12
18	57	50.4	315	2	08ATX7
19	57	50.4	322	2	08NR12
20	57	50.4	363	2	06E072
21	57	50.4	453	2	075C03
22	57	50.4	459	2	09PAU2
23	57	50.4	508	1	MARK_MARSC
24	57	50.4	508	2	09BBG0
25	57	50.4	508	2	09BRG3
26	57	50.4	630	2	08G874
27	57	50.4	1174	2	089NL8
28	57	50.4	1676	2	08TG36
29	56.5	50.0	105	2	06P8X5
30	56.5	50.0	211	2	06ZTC4
31	56.5	50.0	400	2	083A51

32	56	49.6	183	2	06D2A5	Q6d2a5 erwilia car
33	56	49.6	211	2	09N4A6	Q9n4a6 caenorhabdi
34	56	49.6	239	2	06IV35	Q6iv35 photobacter
35	56	49.6	242	2	058743	O58743 schizosacch
36	56	49.6	270	2	09BLU4	Q9blu4 leishmania
37	56	49.6	296	2	08PHM1	Q8phm1 xanthomonas
38	56	49.6	313	2	09N4A7	Q9n4a7 caenorhabdi
39	56	49.6	331	2	067PX1	O67px1 symbiobacte
40	56	49.6	348	2	06IXA4	O6ixa4 methanococc
41	56	49.6	372	2	08MK09	O8mk09 macaca mula
42	56	49.6	372	2	08MK10	O8mk10 macaca mula
43	56	49.6	608	2	0910A3	Q910a3 cyprinus ca
44	56	49.6	2547	1	FAFX_HUMAN	Q93008 h probable
45	55.5	49.1	1749	2	Q8T0W6	Q8t0w6 echinococcu

## ALIGNMENTS

```

RESULT 1
ID 062315 PRELIMINARY; PRT; 107 AA.
AC 062315;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE DCL protein-like.
GN Name=OJ1004_A11.16-2; Synonym=PO539D10.35-2;
OS Oryza sativa (japonica cultivar-group).
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriopharidae; Oryzae; Oryza.
OX NCBI_Taxid=39947;
RN [1]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Katayose Y.;
RL Submitted (MAY-2002) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005287; BAD17315.1; -.
DR EMBL; AP004817; BAD17127.1; -.
SQ SEQUENCE 107 AA; 12839 MW; 00F06AF3A12B0EB1 CRC64;

Query Match 60.2%; Score 68; DB 2; Length 107;
Best Local Similarity 37.5%; Pred. No. 1.3;
Matches 9; Conservative 13; Mismatches 2; Indels 0; Gaps 0;

QY 3 HPQFEKXXXXXXXXXXSHQPE 26
Db 30 HPQYEKKIGGIDYLVGLHPEPE 53

RESULT 2
ID 08CJD2 PRELIMINARY; PRT; 704 AA.
AC 08CJD2;
DT 01-MAR-2003 (TREMBLrel. 23, Created)
DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TREMBLrel. 25, Last annotation update)
DE Guanylyl cyclase alpha 1 subunit.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_Taxid=10116;
RN [1]
RP SEQUENCE FROM N.A.
RA Nakamura T., Suzuki N.;
RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB096020; BAC24016.1; -.
DR HSSP; P30803; IAZS.
GO; GO:0004383; F:guanylate cyclase activity; IEA.

```

DR GO:0016829; F:lyase activity; IEA.  
 DR GO:00007242; P:intracellular signalling cascade; IEA.  
 DR InterPro; IPR001054; G\_cyclase.  
 DR Pfam; PF00211; Guanylate\_cyc; 1.  
 DR SMART; SM00044; CYC; 1.  
 DR PROSITE; PS00452; GUANYLATE\_CYCLASES\_1; 1.  
 DR PROSITE; PS50125; GUANYLATE\_CYCLASES\_2; 1.  
 KW Lyase.  
 SQ SEQUENCE 704 AA; 7918 MW; F73FC97B685446EA CRC64;  
 Query Match 59.3%; Score 67; DB 2; Length 704;  
 Best Local Similarity 45.0%; Pred. No. 15;  
 Matches 9; Conservative 11; Mismatches 0; Indels 0; Gaps 0;

Db 8 KXXXXXXXXXXXXXSHPOFEK 27  
 685 KASGVLDVPRGWSHPQFEK 704

## RESULT 3

ID O8CH90 PRELIMINARY; PRT; 743 AA.  
 AC O8CH90;  
 DT 01-MAR-2003 (TREMBLrel. 23, Created)  
 DT 01-MAR-2003 (TREMBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE Soluble guanylyl cyclase alpha 2 subunit E219G mutant.  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RA Nakamura I., Yao Y., Suzuki N.;  
 RA Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB097860; BAC44887.1; -.  
 DR HSSP; P30803; IAZS.  
 DR GO; GO:0004383; F:guanylate cyclase activity; IEA.  
 DR GO; GO:0016829; F:lyase activity; IEA.  
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.  
 DR InterPro; IPR001054; G\_cyclase.  
 DR InterPro; IPR009080; CRNaseyn\_1a\_bind.  
 DR Pfam; PF00211; Guanylate\_cyc; 1.  
 DR SMART; SM00044; CYC; 1.  
 DR PROSITE; PS00452; GUANYLATE\_CYCLASES\_1; 1.  
 DR PROSITE; PS50125; GUANYLATE\_CYCLASES\_2; 1.  
 KW Lyase.  
 SQ SEQUENCE 743 AA; 83251 MW; ACF5C35E0982813A CRC64;

Query Match 54.9%; Score 62; DB 2; Length 743;  
 Best Local Similarity 42.1%; Pred. No. 78;  
 Matches 8; Conservative 11; Mismatches 0; Indels 0; Gaps 0;

Db 9 XXXXXXXXXXXXXXXXSHPOFEK 27  
 725 LKRTSLVPRGWSHPQFEK 743

## RESULT 4

ID O913C1 PRELIMINARY; PRT; 433 AA.  
 AC O913C1;  
 DT 01-MAR-2001 (TREMBLrel. 16, Created)  
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)  
 DT 01-MAR-2004 (TREMBLrel. 26, Last annotation update)  
 DE Probable cytochrome c.  
 GN OrderedLocustNames=PA1600;  
 OS Pseudomonas aeruginosa.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; Pseudomonas.  
 NCBI\_TaxID=287;  
 RN [1]

RP SEQUENCE FROM N.A.  
 RC STRAIN=ATCC 15692 / PA01;  
 RX MEDLINE=20437337; PubMed=10984043; DOI=10.1038/35023079;  
 RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,  
 RA Hickey M.J., Brinkman F.S.L., Hutnagle W.O., Kowalik D.J., Lagrou M.,  
 RA Garner R.L., Goltzer L., Tolentino E., Westbrock-Madman S., Yuan Y.,  
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,  
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Mu Z., Paulsen I.T.,  
 RA Reizer J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.,  
 RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an  
 RT opportunistic pathogen."  
 RT Nature 406:955-964(2000).  
 RL Nature 406:955-964(2000).  
 DR EMBL; AB004588; AAC04989.1; -.  
 DR PIR; H63444; H63444.  
 DR GO; GO:0004022; F:alcohol dehydrogenase activity; IEA.  
 DR GO; GO:0005489; F:electron transporter activity; IEA.  
 DR GO; GO:0006118; P:electron transport; IEA.  
 DR InterPro; IPR003219; Cytc\_adh.  
 DR InterPro; IPR00345; Cytc\_heme\_BS.  
 DR InterPro; IPR009056; Cytochrome\_c.  
 DR Pfam; PF00034; Cytochrom\_C; 2.  
 DR ProDom; PD01584; Cytc\_adh; 2.  
 DR PROSITE; PS00190; CYTOCHROME\_C; UNKNOWN\_3.  
 KW Complete proteome.  
 SQ SEQUENCE 433 AA; 46248 MW; 83B8E0E75C5E6310 CRC64;

Query Match 54.0%; Score 61; DB 2; Length 433;  
 Best Local Similarity 28.0%; Pred. No. 58;  
 Matches 7; Conservative 15; Mismatches 3; Indels 0; Gaps 0;

Db 1 WSHPOFEKXXXXXXXXXXXXXSHPOF 25  
 99 WSPYAFERAMRHGVARDGSYLTPAF 123

## RESULT 5

ID O8RV8 PRELIMINARY; PRT; 224 AA.  
 AC O8RV8;  
 DT 01-JUN-2002 (TREMBLrel. 21, Created)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last sequence update)  
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)  
 DE DCU protein.  
 GN Name=cl;  
 OS Coffea arabica (Coffee).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; asterids;  
 OC Lamiales; Gentianales; Rubiaceae; Ixoroideae; Coffeae; Coffea.  
 NCBI\_TaxID=13443;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Bean;  
 RA Marrasini P., Meunier A.;  
 RA Submitted (NOV-2001) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AJ420083; CAD12248.1; -.  
 SQ SEQUENCE 224 AA; 25678 MW; E8B58FA50393849E CRC64;

Query Match 53.1%; Score 60; DB 2; Length 224;  
 Best Local Similarity 32.0%; Pred. No. 38;  
 Matches 8; Conservative 13; Mismatches 4; Indels 0; Gaps 0;

Db 3 HPOFEKXXXXXXXXXXXXXSHPOFEK 27  
 151 HPOCEKKGSGVDYITGYHDFDR 175

## RESULT 6

ID O9C642 PRELIMINARY; PRT; 219 AA.  
 AC O9C642;  
 DT 01-JUN-2001 (TREMBLrel. 17, Created)  
 DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)

DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)  
DE Defective chloroplasts and leaves (DCL) protein, putative  
AC (Atg45261/F2G19.1) (Atg45261).  
GN Name=F2G19.1;  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;  
OC eurosid II; Brassicales; Brassicaceae; Arabidopsis.  
OK NCBI\_TaxId=3702;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA Lin X., Raul S., Town C.D., Benito M., Creasy T.H., Haas B.J., Wu D.,  
RA Matl R., Koning C., Koo H., Fujii C.Y., Uteback T.R.,  
RA Barnstead M.E., Bowman C.L., White O., Nierman W.C., Fraser C.M.;  
RL Submitted (JAN-2001) to the EMBL/GenBank/DBJ databases.  
RN (2)  
RP SEQUENCE FROM N.A.  
RA Shin P., Chen H., Cheuk R., Kim C.J., Banh J., Bowser L.,  
RA Carinci P., Chang E., Dale J.M., Goldsmith A.D., Hayashizaki Y.,  
RA Ishida J., Jones T., Kamaya A., Karlin-Neumann G., Kawai J., Lam B.,  
RA Lee J.M., Lin J., Miranda M., Narusaka M., Nguyen M., Onodera C.S.,  
RA Palm C.C., Quach H.L., Sakurai T., Satou M., Seki M., Southwick A.,  
RA Tang C.C., Toriumi M., Wu H.C., Yamada K., Yamamura Y., Yu G., Yu S.,  
RA Shinozaki K., Davis R.W., Theologis A., Ecker J.R.;  
RL Submitted (JUN-2002) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RA Cheuk R., Chen H., Kim C.J., Shin P., Ecker J.R.;  
RL Submitted (JUN-2004) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AC083835; AAG50632.1; -;  
DR EMBL; AY128277; AAM91086.1; -;  
DR EMBL; BT014877; AAT41860.1; -;  
DR PIR; C96510; C96510.  
SQ SEQUENCE 219 AA; 25448 MW; C4061B8DA2A4C448 CRC64;  
Query Match 52.2%; Score 59; DB 2; Length 219;  
Best Local Similarity 33.3%; Pred. No. 51;  
Matches 8; Conservative 12; Mismatches 4; Indels 0; Gaps 0;  
Qy 3 HPOFEKXXXXXXXWSHPQFE 26  
Db 146 HPECEKKGIGIDYIMVGHHPDFE 169  
RESULT 7  
ID Q9AD61 PRELIMINARY; PRT; 389 AA.  
AC Q9AD61;  
DT 01-JUN-2001 (TREMBlrel. 17, Created)  
DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)  
DT 01-JUN-2003 (TREMBlrel. 24, Last annotation update)  
DE Hypothetical protein SCPI.89c.  
DE Ordered locus names=SCPI.89c;  
GN Streptomyces coelicolor.  
OS Streptomyces coelicolor.  
OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
OC Streptomyces; Streptomycetaceae; Streptomyces.  
OK NCBI\_TaxId=1902;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA STRAIN=A3(2) / M145;  
RX MEDLINE=21996410; PubMed=12000953; DOI=10.1038/417141a;  
RA Bentley S.D., Chater K.F., Cerdano-Tarraga A.-M., Challis G.L.,  
RA Thomson N.R., James K.D., Harris D.E., Quail M.A., Kleser H.,  
RA Harper D., Bateman A., Brown S., Chandra G., Chen C.W., Collins M.,  
RA Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S.,  
RA Huang C.-H., Kleser T., Larke L., Murphy L.D., Oliver K., O'Neill S.,  
RA Rabinowitsch E., Rajandream M.A., Rutherford K.M., Rutter S.,  
RA Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K.,  
RA Warren T., Wietzorek A., Woodward J.R., Barrell B.G., Parkhill J.,  
RA Hopwood D.A.;  
RT "Complete genome sequence of the model actinomycete Streptomyces  
coelicolor A3(2)";  
RL Nature 417:141-147(2002).

DR EMBL; AL590463; CAC36610.1; -;  
DR InterPro; IPR009061; Putative DNA bind.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 389 AA; 42800 MW; B27057143383AA27 CRC64;  
Query Match 52.2%; Score 59; DB 2; Length 389;  
Best Local Similarity 25.8%; Pred. No. 97;  
Matches 8; Conservative 15; Mismatches 4; Indels 4; Gaps 1;  
Qy 1 WSHQFEKXXXXXXXXXWS---HPOFEK 27  
Db 266 WAHPEAVTFDGVSRGVGVAHYHPOSDE 296  
RESULT 8  
ID Q62TR3 PRELIMINARY; PRT; 208 AA.  
AC Q62TR3;  
DT 25-OCT-2004 (TREMBlrel. 28, Created)  
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)  
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)  
DE Hypothetical protein.  
GN ORFNames=BL04042;  
OS Bacillus licheniformis DSM 13.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OK NCBI\_TaxId=279010;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA STRAIN=ATCC 14580;  
RA Rey M.W., Ramaiya P., Nelson B.A., Brody-Karpin S.D., Zaretsky E.J.,  
RA Tang M., de Leon A.L., Xiang H., Gusti V., Clausen I.G., Olsen P.B.,  
RA Rasmussen M.D., Andersen J.T., Jorgensen P.L., Larsen T.S.,  
RA Sorokin A., Bolotin A., Lapidus A., Gallon N., Ehrlich S.D.,  
RA Berta R.M.;  
RT "Complete genome sequence of the industrial bacterium Bacillus  
licheniformis and comparisons with closely related Bacillus species";  
RL Genome Biol. 5:R77-R77(2004).  
DR EMBL; CP000002; AAU23846.1; -;  
KW Hypothetical protein.  
SQ SEQUENCE 208 AA; 23430 MW; 20E2053107BD86ED CRC64;  
Query Match 51.8%; Score 58.5; DB 2; Length 208;  
Best Local Similarity 22.2%; Pred. No. 56;  
Matches 8; Conservative 15; Mismatches 4; Indels 9; Gaps 1;  
Qy 1 WSHQFEKXXXXXXXXXWSHPQFEK 27  
Db 153 WAHSEFTLDGALHVALRVARRKRVVLKDWKSPRFK 188  
RESULT 9  
ID Q65IB2 PRELIMINARY; PRT; 256 AA.  
AC Q65IB2;  
DT 25-OCT-2004 (TREMBlrel. 28, Created)  
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)  
DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)  
DE Yp1P.  
GN Name=Yp1P; ORFNames=BL102322;  
OS Bacillus licheniformis DSM 13.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OK NCBI\_TaxId=279010;  
RN (1)  
RP SEQUENCE FROM N.A.  
RA STRAIN=DSM 13;  
RX PubMed=15383718;  
RA Veith B., Herzberg C., Steckel S., Feesche J., Maurer K.H.,  
RA Ehrenreich P., Gottschalk G.,  
RA Ehrenreich P., Baumer S., Henne A., Liesegang H., Mehl R.,  
RT "The Complete Genome Sequence of Bacillus licheniformis DSM13, an  
Organism with Great Industrial Potential";  
RL J. Mol. Microbiol. Biotechnol. 7:204-211(2004).  
DR EMBL; AE017333; AAU41202.1; -;

```

SQ      SEQUENCE   256 AA;  29029 MW;  C127FF81C043FBFAB9 CRC64;

Query Match           51.8%; Score 58.5; DB 2; Length 256;
Best Local Similarity 22.4%; Pred. No. 71;
Matches      8; Conservative 15; Mismatches    4; Indels    9; Gaps    1;

Qy      1 WSHPOFE-----KXXXXXXXXXXWSHPOFEK 27
        |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db       201 WAHSEFTLDGALHEALVARAKRVVLDHKMSRPFEX 236

RESULT 10
ID      O8X549          PRELIMINARY; PRT; 103 AA.
AC      O8X549; Q7ADVO;
DT      01-MAR-2002 (TrEMBLrel. 20, Created)
DI      01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE      Hypothetical protein Z2097 (Hypothetical protein ECs2201).
DS      OrderedLOCusNames=ECs2201, z2097;
GN      Escherichia coli O157:H7.
OC      Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OX      Enterobacteriaceae; Escherichia.
NCBI    NCBI_Taxid=83334;
PP      [1] SEQUENCE FROM N.A.
RC      STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;
RX      MEDLINE=21074935; PubMed=11206551; DOI=10.1038/35054089;
RA      Perna N.T., Plunkett G. III, Burland V., Mau B., Claener J.D.,
RA      Rose D.J., Mayhew G.F., Evans P.S., Gregor U., Kirkpatrick H.A.,
RA      Postel G., Hackert J., Klink S., Boutin A., Shao Y., Miller L.,
RA      Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamoustis K.,
RA      Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA      Welch R.A., Blattner F.R.;
RT      "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";
RL      Nature 409:529-533(2001).
[2]
RP      SEQUENCE FROM N.A.
RC      STRAIN=O157:H7 / RIND 0509952 / EHEC;
RX      MEDLINE=2115831; PubMed=11258796;
RA      Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA      Han C.-G., Okubo E., Nakayama K., Murata T., Tanaka M., Toke T.,
RA      Iida T., Takami H., Honda T., Saekawa C., Ogatawara N., Yasunaga T.,
RA      Kubura S., Shibata T., Hattori M., Shinagawa H.;
RT      "Complete genome sequence of enterohemorrhagic Escherichia coli
RT      O157:H7 and genomic comparison with a laboratory strain K-12.";
RL      DNA Res. 8:11-22(2001).
DR      EMBL; AE005346; AAC56167.1; -
DR      EMBL; AP002557; BAB35624.1; -.
DR      PIR; A90904; A90904.
DR      PIR; C85713; C85713.
KW      Hypothetical protein; Complete proteome.
SQ      SEQUENCE 103 AA; 11980 MW; 5A42AAFA91CF29EB6 CRC64;

Query Match           51.3%; Score 58; DB 2; Length 103;
Best Local Similarity 30.0%; Pred. No. 30;
Matches      6; Conservative 14; Mismatches    0; Indels    0; Gaps    0;

Qy      8 KXXXXXXXXXXXXWSHPOFEK 27
        |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
Db       5 KEMPVVRDYGWYTHPEYEX 24

RESULT 11
ID      O8X4VO          PRELIMINARY; PRT; 103 AA.
AC      O8X4VO; Q7ACU2;
DT      01-MAR-2002 (TrEMBLrel. 20, Created)
DI      01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
DE      DT 25-OCT-2004 (TrEMBLrel. 28, Last annotation update)
DS      Hypothetical protein Z3120 (Hypothetical protein ECs2757).
GN      OrderedLOCusNames=ECS2757, z3120;
OC      Escherichia coli O157:H7.
```

```

OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
OC Enterobacteriaceae; Escherichia.
OX NCBI_TaxID=83334;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=0157:H7 / EDL933 / ATCC 700927 / EHEC;
RC MEDLINE=21074935 / PubMed=11206551; DOI=10.1038/35054069;
RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,
RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller E.,
RA Grodeck J.E., Davis N.W., Lim A., Dimalanta E.T., Potamovasis K.,
RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
RA Welch R.A., Blattner F.R.;
RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7,"
RL Nature 409:529-533 (2001).
RP
RN
RP SEQUENCE FROM N.A.
RC STRAIN=0157:H7 / RIMD 0509952 / EHEC;
RC MEDLINE=21156231 / PubMed=11258796;
RA Hayashi T., Makino K., Onishi M., Kurokawa K., Ishii K., Yokoyama K.,
RA Han C.-o., Okubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,
RA Iida T., Takami H., Honda T., Saekawa C., Ogasawara N., Yasunaga T.,
RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
RT O157:H7 and genomic comparison with a laboratory strain K-12."
RL DNA Res. 8:11-22(2001).
DR EMBL; AF005422; AAG57029.1; -.
DR EMBL; AP002559; BAB36180.1; -.
DR PIR; A65821; A65821.
DR PIR; E90973; E90973.
KW Hypothetical protein; Complete proteome.
SQ
SEQUENCE 103 Aa; 12092 MW; DA43CEA5A1089D30 CRC64;

Query Match 51.3%; Score 58; DB 2; Length 103;
Best Local Similarity 30.0%; Pred. NO. 30;
Matches 6; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

Qy 8 KXXXXXXXXXXMHPQPEK 27
|:::|||||:|::|
Db 5 KEMPERNEYGCMTHPEYK 24

RESULT 12
Q88MQ4 PRELIMINARY; PRT; 366 AA.
AC Q88MQ4.
DT 01-JUN-2003 (TRENBLrel. 24, Created)
DT 01-JUN-2003 (TRENBLrel. 24, Last sequence update)
DT 01-MAR-2004 (TRENBLrel. 26, Last annotation update)
DE RND membrane fusion protein.
GN OrderedLocustNames=PI1516;
OS Pseudomonas putida (strain KT2440).
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;
OC Pseudomonadaceae; Pseudomonas.
NCBI_TaxID=160488;
[1]
RN
RP SEQUENCE FROM N.A.
RC MEDLINE=2423060; PubMed=12534463;
RA Nelson K.E., Weinl C., Paulsen I.T., Dodson R.J., Hilbert H.,
RA Martins dos Santos V.A.P., Fouts D.B., Gill S.R., Pop M., Holmes M.,
RA Brinkac L.M., Beanan M.J., DeBoy R.T., Daugherty S.C., Kolonay J.F.,
RA Hancu I., Nelson W.C., White O., Peterson J.D., Khouri H.M.,
RA Madupu R., Chris Lee P., Holtzapple E.K., Scanlan D., Tran K.,
RA Moazzaz A., Ullrich T.R., Rizzo M., Lee K., Kosack D., Woese D.,
RA Wedler H., Hauber J., Stepanick D., Hohnesiel J., Straetz M., Heim S.,
RA Kiewitz C., Eisen J.A., Timmis K.N., Duesterhoeft A., Tuenmler B.,
RA Fraser C.M.;
RT "Complete genome sequence and comparative analysis of the
RT metabolically versatile Pseudomonas putida KT2440."
RL Environ. Microbiol. 4:799-808 (2002).
DR EMBL; AE016779; AAN67137.1; -.
DR TIGR; PP1516; -.
GO; GO:0016020; C:membrane; IEA.

```



CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; M60052; AAA88071.1; -

DR PIR; A54660; A54660.

DR Genew; HGNC:5178; HRC.

DR MIM; 142705; -

DR GO; GO:0005509; F:calcium ion binding; TAS.

DR GO; GO:006936; P:muscle contraction; TAS.

KW Calcium-binding; Polymorphism; Repeat; Signal.

FT SIGNAL 1 28

FT CHAIN 29 699

FT DOMAIN 193 204

FT DOMAIN 246 261

FT DOMAIN 106 365

FT REPEAT 180 213

FT REPEAT 238 270

FT REPEAT 295 318

FT REPEAT 343 365

FT DOMAIN 106 342

FT REPEAT 106 121

FT REPEAT 134 154

FT REPEAT 155 177

FT REPEAT 214 237

FT REPEAT 271 294

FT REPEAT 319 342

FT DOMAIN 627 673

FT VARIANT 96 96

FT VARIANT 204 204

FT SEQUENCE 699 AA; 80244 MW; 9922EEDF012C61DD CRC64;

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

FT

Search completed: March 2, 2005, 12:44:30  
 Job time : 101.317 secs

QY 3 HPOFEKXXXXXXXXXXXXXGHPQFEK 27  
 DB 67 HPD-EKKDVSTENGHHFWSPDREK 90

Query Match 50.9%; Score 57.5; DB 1; Length 699;  
 Best Local Similarity 36.0%; Pred. No. 3e+02;  
 Matches 9; Conservative 11; Mismatches 4; Indels 1; Gaps 1;

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 121.61 Seconds  
(without alignment)  
85.869 Million cell updates/sec

Title: SEQ15  
Perfect score: 113  
Sequence: 1 wshpqfexkxxxxxxxxxxwshpqfex 27

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq15dec04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	99	87.6	35	6	ABP60369 Abp60369 Streptavi
2	98.5	87.2	24	6	ABP60362 Abp60362 Streptavi
3	98.5	87.2	36	6	ABP60370 Abp60370 Streptavi
4	69.5	61.5	24	6	ABP60363 Abp60363 Streptavi
5	67.5	59.7	646	8	ADS20251 Ades0251 Human ags
6	67.5	59.7	858	8	ADS20230 Ades0230 Human ags
7	67	59.3	490	4	AAM40759 Aam40759 Human pol
8	66	58.4	485	8	ADS20249 Ades20249 Furin-pro
9	66	58.4	697	8	ADS20227 Ades20227 Human ags
10	63.5	56.2	179	5	AAB24889 Aab24889 Pneumococ
11	63	55.8	448	4	AAB46425 Aab46425 Bilin bin
12	62.5	55.3	236	8	ADO19051 Ado19051 Murine an
13	62.5	55.3	236	8	ADO19053 Ado19053 Murine an
14	62	54.9	117	5	AAU97558 Aau97558 Synthetic
15	62	54.9	117	5	AAU97553 Aau97553 Synthetic
16	62	54.9	117	5	AAU97557 Aau97557 Synthetic
17	62	54.9	117	5	AAU97559 Aau97559 Synthetic
18	62	54.9	117	5	AAU97555 Aau97555 Synthetic
19	62	54.9	117	5	AAU97556 Aau97556 Synthetic
20	62	54.9	117	5	AAU97552 Aau97552 Synthetic
21	62	54.9	117	5	AAU97560 Aau97560 Synthetic
22	62	54.9	118	5	AAU97554 Aau97554 Synthetic
23	62	54.9	183	6	ADA27294 Adas27294 Plasmid p
24	62	54.9	183	6	ADA00703 Ada00703 Modified
25	62	54.9	199	7	ADB87309 Adb87309 Apolipop

26	54.9	205	2	AAW93967 Aaw93967 Plasmid p
27	62	209	6	ADA00709 Ada00709 Modified
28	54.9	248	7	ADB87310 Adb87310 Apolipop
29	62	254	2	AAW93969 Aaw93969 Plasmid p
30	62	254	4	AAB46421 Aab46421 Bilin bin
31	62	258	6	ADA27293 Adas27293 Plasmid p
32	54.9	258	6	ADA00702 Ada00702 Modified
33	54.9	267	8	ADR70320 Adr70320 Polioviru
34	62	268	8	ADR70385 Adr70385 Polioviru
35	62	268	8	ADR70319 Adr70319 Polioviru
36	62	268	8	ADR70386 Adr70386 Polioviru
37	62	269	8	ADR28054 Adr28054 NPB polyP
38	54.9	269	8	ADS17525 Ades17525 Amino aci
39	62	273	8	ADO25157 Ados25157 Melanoma
40	62	274	8	ADO25153 Ados25153 Melanoma
41	62	275	8	ADO25154 Ados25154 Melanoma
42	62	275	8	ADO25155 Ados25155 Melanoma
43	62	276	8	ADO25149 Ados25149 Melanoma
44	62	277	8	ADO25150 Ados25150 Melanoma
45	62	277	8	ADO25152 Ados25152 Melanoma

## ALIGNMENTS

RESULT 1	ABP60369	ABP60369 standard; peptide; 35 AA.
ID	XX	XX
AC	XX	ABP60369;
DT	XX	28-MAR-2003 (first entry)
DE	XX	Streptavidin binding peptide SEQ ID NO 10.
XX	XX	Streptavidin; protein chip; microtitre plate; detection.
KW	XX	Synthetic.
OS	XX	
FH	XX	Key Location/Qualifiers
FT	XX	Misc-difference 9..27
FT	XX	/label= unknown
FT	XX	/note= "optionally deleted for 1-15 residues"
XX	XX	
PN	XX	DE10113776-AL.
PD	XX	02-OCT-2002.
PF	XX	21-MAR-2001; 2001DE-01013776.
PR	XX	21-MAR-2001; 2001DE-01013776.
PA	XX	(BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
PI	XX	Schmidt T;
DR	XX	WPI; 2003-031166/03.
PT	XX	New isolated peptide, useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.
XX	XX	Claim 7; Page 16; 18pp; German.
PS	XX	The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides stronger binding than a single tag, but are displaced by a competitor. (I) does not interfere with the function of attached proteins (II) (so it

CC may not be essential to remove it); facilitates detection and has easily  
CC controllable binding properties. (I) is particularly used for purifying  
CC FP from dilute solution in batch formats (which use simpler apparatus  
CC than column methods and result in lower loss of FP). The present sequence  
CC is that of a streptavidin binding peptide disclosed with the invention  
XX  
SQ Sequence 35 AA;

Query Match 87.6%; Score 99; DB 6; Length 35;  
Best Local Similarity 77.1%; Pred. No. 8.8e-06;  
Matches 27; Conservative 0; Mismatches 0; Indels 8; Gaps 1;

QY 1 WSHPOFEK-----XXXXXXXXXXXXXWSHPOFEK 27  
DB 1 WSHPOFEKXXXXXXXXXXXXXXXXXXXXXWSHPOFEK 35

## RESULT 2

ABP60362 ID ABP60362 standard; peptide; 24 AA.

AC ABP60362;

DT 28-MAR-2003 (first entry)

DE Streptavidin binding peptide SEQ ID NO 3.

XX Streptavidin; protein chip; microtitre plate; detection.

OS Synthetic.

PH Key Location/Qualifiers

FT Misc-difference 9.16

FT /label= unknown

PN DE10113776-A1.

PD 02-OCT-2002.

PF 21-MAR-2001; 2001DE-01013776.

PR 21-MAR-2001; 2001DE-01013776.

PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

PI Schmidt T;

DR WPI; 2003-031166/03.

PT New isolated peptide, useful as affinity purification tag for recombinant  
PT protein, comprises at least two high-affinity streptavidin-binding  
PT modules.

PS Disclosure; Page 4; 18pp; German.

CC The invention relates to an isolated peptide (I) comprising at least two  
CC individual modules separated by 0-50 amino acids, with each containing at  
CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain  
CC streptavidin binding modules, are useful as affinity handles for  
CC purification of recombinant fusion proteins (FP), also for detecting FP,  
CC e.g. on protein chips or microtitre plates. The modules in (I) bind  
CC strongly to streptavidin, with a co-operative effect that provides  
CC stronger binding than a single tag, but are displaced by a competitor.  
CC (I) does not interfere with the function of attached proteins (II) (so it  
CC may not be essential to remove it); facilitates detection and has easily  
CC controllable binding properties. (I) is particularly used for purifying  
CC FP from dilute solution in batch formats (which use simpler apparatus  
CC than column methods and result in lower loss of FP). The present sequence  
CC is that of a streptavidin binding peptide disclosed with the invention  
XX  
SQ Sequence 24 AA;

Query Match 87.2%; Score 98.5; DB 6; Length 24;

Best Local Similarity 88.9%; Pred. No. 5.9e-06;  
Matches 24; Conservative 0; Mismatches 0; Indels 3; Gaps 1;  
QY 1 WSHPOFEKXXXXXXXXXXXXXWSHPOFEK 27  
DB 1 WSHPOFEK-----XXXXXXXXXXXXXWSHPOFEK 24

## RESULT 3

ABP60370 ID ABP60370 standard; peptide; 36 AA.

AC ABP60370;

DT 28-MAR-2003 (first entry)

DE Streptavidin binding peptide SEQ ID NO 11.

XX Streptavidin; protein chip; microtitre plate; detection.

OS Synthetic.

PH Key Location/Qualifiers

FT Region

FT /note= "GGGS repeats 2-5 optionally absent, residues 13-  
28"

FT Region 9.12

FT /label= GGGS\_repeat

PN DE10113776-A1.

PD 02-OCT-2002.

PF 21-MAR-2001; 2001DE-01013776.

PR 21-MAR-2001; 2001DE-01013776.

PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

PI Schmidt T;

DR WPI; 2003-031166/03.

PT New isolated peptide, useful as affinity purification tag for recombinant  
PT protein, comprises at least two high-affinity streptavidin-binding  
PT modules.

PS Claim 8; Page 16; 18pp; German.

CC The invention relates to an isolated peptide (I) comprising at least two  
CC individual modules separated by 0-50 amino acids, with each containing at  
CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain  
CC streptavidin binding modules, are useful as affinity handles for  
CC purification of recombinant fusion proteins (FP), also for detecting FP,  
CC e.g. on protein chips or microtitre plates. The modules in (I) bind  
CC strongly to streptavidin, with a co-operative effect that provides  
CC stronger binding than a single tag, but are displaced by a competitor.  
CC (I) does not interfere with the function of attached proteins (II) (so it  
CC may not be essential to remove it); facilitates detection and has easily  
CC controllable binding properties. (I) is particularly used for purifying  
CC FP from dilute solution in batch formats (which use simpler apparatus  
CC than column methods and result in lower loss of FP). The present sequence  
CC is that of a streptavidin binding peptide disclosed with the invention  
XX  
SQ Sequence 36 AA;

Query Match 87.2%; Score 98.5; DB 6; Length 36;  
Best Local Similarity 44.4%; Pred. No. 1.1e-05;  
Matches 16; Conservative 11; Mismatches 0; Indels 9; Gaps 1;

QY 1 WSHPOFEK-----XXXXXXXXXXXXXWSHPOFEK 27  
DB 1 WSHPOFEKGGGGGGGGGGGGGGGGGGGGGGGGWSHPOFEK 36



ID	ABP60363	standard, peptide; 24 AA.
XX	ABP60363;	
XX	28-MAR-2003 (first entry)	
XX	Streptavidin binding peptide SEQ ID NO 4.	
XX	Streptavidin, protein chip, microtitre plate, detection.	
XX	Synthetic.	
XX	Key	Location/Qualifiers
XX	Misc-difference 9..18	/label= unknown
XX	Misc-difference 22..24	/label= unknown
XX	DE10113376-A1.	
XX	02-OCT-2002.	
XX	21-MAR-2001; 2001DE-010133776.	
XX	21-MAR-2001; 2001DE-010133776.	
XX	(BIOA-) INST BIOMALYTIK GMBH GOETTINGEN.	
XX	Schmidt T;	
XX	WPI; 2003-031166/03.	
XX	New isolated peptide, useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.	
XX	Disclosure; Page 4; 18pp; German.	
XX	The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides stronger binding than a single tag, but are displaced by a competitor. (I) does not interfere with the function of attached proteins (II) (so it may not be essential to remove it); facilitates detection and has easily controllable binding properties. (I) is particularly used for purifying FP from dilute solution in batch formats (which use simpler apparatus than column methods and result in lower loss of FP). The present sequence is that of a streptavidin binding peptide disclosed with the invention	
XX	Sequence 24 AA:	
XX	Query Match	61.5%; Score 69.5; DB 6; Length 24;
XX	Best Local Similarity	87.5%; Pred. No. 0.07;
XX	Matches 21; Conservative	0; Mismatches 0; Indels 3; Gaps 1
XX	1 WSHPOPEKXXXXXXXXXXWSHPQ 24	
XX	1 WSHPOPEKXXXXXXXXXXWSHPQ 21	

```

XX 18-NOV-2004      (first entry)
DT
XX
XX Human aggrecanase modified ADAMTS4 (MTS4) protein - SEQ ID 49.
DE
XX ADAMTS4; a disintegrin-like and metalloprotease;
KW Chondrospondin type I motif 4; reprotysin; zinc metalloprotease;
KW aggrecanase; osteoarthritis; antiinflammatory; antirheumatic;
KW cystostatic; osteoarthritis; glioma; cancer; inflammatory joint;
KW rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;
KW human; chromosome 1q21-q23; enzyme.
XX Homo sapiens.
OS Synthetic.
XX
XX WO2004011637-A2.
XX
XX 05-FEB-2004.
PD
XX
XX 29-JUL-2003; 2003WO-US023484.
PF
XX
XX 29-JUL-2002; 2002US-0398721P.
PR
XX
XX (AMHP ) WYETH.
PA (CORC/) CORCORAN C J.
PA (PLAN/) FLANNERY C R.
PA (ZENG/) ZENG W.
PA (RACI/) RACIE L A.
PA (MCDON/) MCDONAGH T.
PA (FREE/) FREEMAN B A.
PA (GEOR/) GEORGIADIS K E.
PA (LAVA/) LAVALLIE E R.
XX
XX Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T;
PI Freeman BA, Georgiadis KE, Lavallie ER;
PI
XX WPI; 2004-143860/14.
XX
XX New isolated, modified ADAMTS4 (aggrecanase) protein with improved
PT stability useful for identifying inhibitors of the enzyme activity for
PT treating aggrecanase-associated conditions, including osteoarthritis.
PT
XX Claim 9; SEQ ID NO 49; 117bp; English.
PS
XX
XX The invention relates to a novel isolated, modified ADAMTS4 (a
XX chondrospondin-like and metalloprotease (reprotysin type) with
XX chromospondin type I motif 4) protein with improved stability compared
CC to a naturally occurring, full-length ADAMTS4 protein, where the modified
CC protein differs from the naturally-occurring, full-length ADAMTS4 protein
CC by at least one amino acid. ADAMTS proteins are a subfamily of zinc
CC metalloproteinases and include aggrecanases amongst their members. The
CC protein of the invention demonstrates osteopathic, antiinflammatory,
CC antiarthritic, antirheumatic and cytostatic activities and may be useful
CC for treating aggrecanase-associated conditions, including osteoarthritis,
CC glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic
CC arthritis, periodontal disease and Crohn's disease. The current sequence
CC is that of a human aggrecanase modified ADAMTS4 (mts4) protein of the
CC invention.
XX
XX Sequence 646 AA;
SQ
XX
Query Match          59.7%; Score 67.5; DB 8; Length 646;
Best Local Similarity 31.2%; Pred. No. 17;
Matches 10; Conservative 14; Mismatches 3; Indels 5; Gaps 1
Qy 1 WSHPOFE-----KXXXXXXXXXXXXWSHPQFEX 27
   |||:::|||||||
Db 615 WLHRRRAQLLEILRRPPWAGRGSAMWSHPQFEX 646

```

XX AC ADS20230;  
 XX XX 18-NOV-2004 (first entry)  
 DT XX  
 DE Human aggrecanase ADAMTS4 truncated protein with insert/Strep tag.  
 KW ADAMTS4, a disintegrin-like and metalloprotease;  
 KW thrombospondin type 1 motif 4; reprolysin; zinc metalloprotease;  
 KW cytosolic; osteopathic; antiinflammatory; zinc metalloprotease;  
 KW rheumatoid arthritis; glioma; cancer; inflammatory joint; antirheumatic;  
 KW human; enzyme; chromosome 12q1-q23; truncation; Strep tag; mutant;  
 XX XX  
 OS Homo sapiens.  
 OS Synthetic.  
 OS Unidentified.  
 PH Key  
 FT Misc-difference 430.439  
 FT Location/Qualifiers  
 FT /note="Flexible spacer/enterokinase recognition site  
 FT Misc-difference 848.858  
 FT /note="Strep tag"  
 PN WO2004011337-A2.  
 XX PD 05-FEB-2004.  
 XX PF 29-JUL-2003; 2003WO-US023484.  
 XX PR 29-JUL-2002; 2002US-0398721P.  
 XX PA (AMBP) WYETH.  
 XX PA (CORC) CORCORAN C J.  
 XX PA (ZENG) ZENG W.  
 XX PA (RAC) RACIE L A.  
 XX PA (MCD) MCDONAGH T.  
 XX PA (PRE) FREEMAN B A.  
 XX PA (GEOR) GEORGIADIS K E.  
 XX PA (LAVA) LAVALLIE E R.  
 PI Corcoran C J, Flannery CR, Zeng W, Racie LA, McDonagh T,  
 PI Freeman BA, Georgiadis KE, Lavallie ER, McDonagh T;  
 DR WPI; 2004-143860/14.  
 PT New isolated, modified ADAMTS4 (aggrecanase) protein with improved  
 PT stability useful for identifying inhibitors of the enzyme activity for  
 PT creating aggrecanase-associated conditions, including osteoarthritis.  
 PS Claim 9; SEQ ID NO 27; 117bp; English.  
 CC The invention relates to a novel isolated, modified ADAMTS4 (a  
 CC disintegrin-like and metalloprotease (reprolysin type) with  
 CC thrombospondin type 1 motif 4) protein with improved stability  
 CC to a naturally occurring, full-length ADAMTS4 protein, where the modified  
 CC by at least one amino acid ADAMTS4 protein, where the modified  
 CC metalloprotease and include aggrecanase, full-length ADAMTS4 protein  
 CC protein of the invention demonstrates osteopathic, antiinflammatory,  
 CC antirheumatic, antineoplastic and cytostatic activities, their members. The  
 CC for treating aggrecanase-associated conditions, including osteoarthritis,  
 CC glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic  
 CC arthritis, periodontal disease and Crohn's disease. The current sequence  
 CC is that of the human aggrecanase ADAMTS4 truncated protein with insert  
 CC peptide and Strep tag of the invention.  
 XX Sequence 858 AA;  
 SQ  
 Query Match 59.7%; Score 67.5; DB 8; Length 858;

## seq15.rag

Best Local Similarity 31.2%, Pred. No. 26;  
 Matches 10; Conservative 14; Mismatches 3; Indels 5; Gaps 1.  
 QY 1 MSHPQPE-----KXXXXXXXXXXMSHPQPEK 27  
 DB 827 MHRRAQILRLRRRRWRGRKGSAMSHPQPEK 858  
 RESULT 7  
 ID AAM40759  
 AC AAM40759 standard; protein, 490 AA.  
 XX AAM40759;  
 DT 22-OCT-2001 (first entry)  
 DE Human polypeptide SEQ ID NO 5690.  
 KW Human; neotropic; immunosuppressant; cytostatic; gene therapy; cancer;  
 KW peripheral nervous system; neuropathy; central nervous system; CNS;  
 KW Alzheimer's; Parkinson's disease; Huntington's disease; haemostatic;  
 KW chemokine; thrombolytic; drug screening; arthritis; inflammation;  
 XX leukemia.  
 OS Homo sapiens.  
 XX XX  
 PN WO200153312-A1.  
 XX PD 26-JUL-2001.  
 XX PF 26-DEC-2000; 2000WO-US034263.  
 XX PR 23-DEC-1999; 99US-00471275.  
 XX PR 21-JAN-2000; 2000US-0048725.  
 XX PR 20-APR-2000; 2000US-00552317.  
 XX PR 19-JUN-2000; 2000US-0058042.  
 XX PR 03-AUG-2000; 2000US-00620312.  
 XX PR 14-SEP-2000; 2000US-00653450.  
 XX PR 19-OCT-2000; 2000US-00662191.  
 XX PR 29-NOV-2000; 2000US-00693036.  
 XX PA (HYSB-) HYSBQ INC.  
 PI Tang YT, Liu C, Aundi V, Chen R, Ma Y, Qian XB, Ren F, Wang D;  
 PI Wang J, Wang Z, Wehrman T, Xu C, Xue MJ, Yang Y, Zhang J, Zhao Q;  
 PI Zhou P, Goodrich R, Drmanac RT;  
 DR WPI; 2001-442253/47.  
 XX N-PSDB; AAI59915.  
 PT Novel nucleic acids and polypeptides, useful for treating disorders such  
 PT as central nervous system injuries.  
 PS Example 2; SEQ ID NO 5690; 10078bp; English.  
 CC The invention relates to human nucleic acids (AAI5798-AAI1369) and the  
 CC encoded polypeptides (AAM38642-AAM42213) with neotropic,  
 CC immunosuppressant and cytostatic activity. The polynuc.  
 CC in gene therapy. A composition containing a polypeptide or polynucleotide  
 CC of the invention may be used to treat diseases of the peripheral nervous  
 CC system, such as peripheral nervous injuries, peripheral neuropathy and  
 CC Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
 CC lateral sclerosis, and Shy-Drager Syndrome. Other uses include the  
 CC utilization of the activities such as: immune system suppression,  
 CC activation/inhibition activity, chemokine/chemokine activity, haemostatic  
 CC and thrombolytic activity, cancer diagnosis and therapy, drug screening,  
 CC assays for receptor activity, arthritis and inflammation, leukemias and  
 CC C.N.S disorders. Note: The sequence data for this patent did not form  
 CC part of the printed specification

Seq	Sequence	490 AA;
Query Match	59.3%; Score 67; DB 4; Length 490;	
Best Local Similarity	34.6%; Pred. No. 13;	
Matches	9; Conservative 13; Mismatches 4; Indels 0; Gaps 0;	
Qy	2 SHIPEKXXXXXXXXXXWSHPQEK 27	
Db	394 SHPARKSTSMSEAKSYWNPHRDR 419	
RESULT 8		
ADSS20249		
ID	ADSS20249 standard; protein; 485 AA.	
AC	ADSS20249;	
XX		
DT	18-NOV-2004 (first entry)	
DE	Furin-processed human aggrecanase ADAMTS4 truncated protein w Streptag.	
XX		
XX		
KW	ADAMTS4; a disintegrin-like and metalloprotease;	
KW	chondropondin type 1 motif 4; reprotolysin; zinc metalloprotease;	
KW	aggrecanase; osteoarthritis; antiinflammatory; antirheumatic;	
KW	cytostatic; osteoarthritis; glioma; cancer; inflammatory joint;	
KW	rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;	
KW	human; chromosome 1q21-q23; enzyme; truncation; mature; furin cleavage;	
XX	Streptag.	
XX		
OS	Homo sapiens.	
OS	Unidentified.	
XX		
FT	Key Location/Qualifiers	
FT	Misc-difference 475..485	
FT	/note="Residues corresponding to positions 687-837 in	
FT	the wild-type replaced by Streptag"	
XX		
XX	WO2004011637-A2.	
XX		
PD	05-FEB-2004.	
XX		
PF	29-JUL-2003; 2003WO-US023484.	
XX		
PR	29-JUL-2002; 2002US-0398721P.	
XX		
PA	(AAMP) WYETH.	
PA	(CORC) CORCORAN C J.	
PA	(FLAN)/ FLANNERY C R.	
PA	(ZENG)/ ZENG W.	
PA	(RACI)/ RACIE L A.	
PA	(MCDON)/ MCDONAGH T.	
PA	(FREE)/ FREEMAN B A.	
PA	(GEOR)/ GEORGIADIS K E.	
PA	(LAVA)/ LAVALLIE E R.	
XX		
PI	Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T;	
PI	Freeman BA, Georgiadis KE, Lavallie ER;	
XX		
XX	WPI, 2004-143860/14.	
DR		
XX		
XX		
PT	New isolated, modified ADAMTS4 (aggrecanase) protein with improved	
PT	stability useful for identifying inhibitors of the enzyme activity for	
PT	treating aggrecanase-associated conditions, including osteoarthritis.	
XX		
XX	Claim 9; SEQ ID NO 47; 117pp; English.	
XX		
CC	The invention relates to a novel isolated, modified ADAMTS4 (a	
CC	disintegrin-like and metalloprotease (reprotolysin type) with	
CC	chondropondin type 1 motif 4) protein with improved stability compared	
CC	to a naturally occurring, full-length ADAMTS4 protein, where the modified	
CC	protein differs from the naturally-occurring, full-length ADAMTS4 protein	
CC	by at least one amino acid. ADAMTS proteins are a subfamily of zinc	
CC	metalloproteases and include aggrecanases amongst their members. The	

CC		protein of the invention demonstrates osteopathic, antiinflammatory,
CC		antiarthritic, antithumatic and cytostatic activities and may be useful
CC		for treating aggrecanase-associated conditions, including osteoarthritis,
CC		glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic
CC		arthritis, periodontal disease and Crohn's disease. The current sequence
CC		is that of the furin-processed human aggrecanase ADAMTS4 truncated
XX		protein with Strep tag of the invention.
SQ	Sequence 485 AA;	
Query Match	58.4%; Score 66; DB 8; Length 485;	
Best Local Similarity	36.7%; Pred. NO. 18;	
Matches 11; Conservative 13; Mismatches 2; Indels 4; Gaps 1		
CY	2 SHPQEK-----XXXXXXXXXXXXXWSHPQEK 27   :   :   :::::	
Dd	456 SKKRFCKMVGCGGDSGCCSGSMWSPQEK 485	
RESULT 9		
ADSS20227		
ID	ADS20227 standard; protein; 697 AA.	
AC	ADS20227;	
DT	18-NOV-2004 (first entry)	
DE	Human aggrecanase ADAMTS4 truncated protein with Strep tag - SEQ ID 24.	
KX	ADAMTS4; a disintegrin-like and metalloprotease;	
KM	chondrospondin type I motif 4; reprolysin; zinc metalloprotease;	
KM	aggrecanase; osteoblastic; antiinflammatory; antiarthritic; antirheumatic;	
KW	cystostic; osteoarthritis; glioma; cancer; inflammatory joint;	
KV	rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;	
KW	human; enzyme; chromosome 1q21-q23; truncation; Strep tag.	
OS	Homo sapiens.	
OS	Unidentified.	
FH	Key Location/Qualifiers	
FT	Misc-difference 687..697 /note="Wild-type residues 687-837 replaced by Strep tag"	
PN	WO2004011637-A2.	
PD	05-FEB-2004.	
PF	29-JUL-2003; 2003WO-US023484.	
PR	29-JUL-2002; 2002US-0398721P.	
PA	(AMRP ) WYETH-	
PA	(CORC/) CORCORAN C J.	
PA	(FLAN/) FLANNERY C R.	
PA	(ZENG/) ZENG W.	
PA	(RACI/) RACIE L A.	
PA	(MCDO/) McDONAGH T.	
PA	(FREE/) FREEMAN B A.	
PA	(GEOR/) GEORGIAIDS K E.	
PA	(LAVA/) LAVALLIE E R.	
PI	Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T;	
PI	Freeman BA, Georgiadis KE, Lavallie ER,	
DR	WPI; 2004--143860/14.	
New isolated, modified ADAMTS4 (aggrecanase) protein with improved stability useful for identifying inhibitors of the enzyme activity for treating aggrecanase-associated conditions, including osteoarthritis.		
Claim 9; SBQ ID NO 24; 117pp; English.		
The invention relates to a novel isolated, modified ADAMTS4 (a		



```

PF 08-JUN-2000; 2000WO-DE001873.
XX
PR 08-JUN-1999; 99DE-01026068.
XX
PA (SKER/) SKERRA A.
XX
PI Skerra A, Schlehuber S;
XX
DR MPI; 2001-071071/08.
DR N-PSDB; AAF25710.
XX
PT New muteins of bilin-binding protein, useful for detecting digoxigenin
PT being used as label in e.g. binding assays, are very selective for
PT digoxigenin.
XX
PS Example 3; Page 65-67; 80pp; German.
XX
CC This invention describes novel polypeptides (I) that are muteins of bilin
CC -binding protein (BBP), which can bind digoxigenin (Dig) or its
CC conjugates, do not bind ouabain, testosterone or 4-aminofluorescein, and
CC have an amino acid (aa) substitution at at least one of the positions 28,
CC 31, 34-37, 58, 60, 69, 88, 90, 95, 97, 114, 116, 125 or 127. (I), or its
CC fusion proteins, are used to bind, detect, determine, immobilize or
CC separate Dig or its conjugates with proteins, nucleic acids,
CC carbohydrates, other biological or synthetic macromolecules or low
CC molecular weight compounds, particularly in assays where Dig is being
CC used as a label. Compared with Dig-specific antibodies, (I) have a
CC simpler structure and are easier to prepare. They have very high
CC specificity for Dig, relative to other steroids, and fusion partners may
CC be attached to either end without compromising their ability to bind
CC ligand
XX
SQ Sequence 448 AA;

Query Match          55.8%; Score 63; DB 4; Length 448;
Best Local Similarity 50.0%; Pred. No. 42;
Matches 10; Conservative 10; Mismatches 0; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXXW 20
   |||||
DB 198 WSHPOFEKXXASGRFWMKGF 217

RESULT 12
AD019051
ID AD019051 standard; protein; 236 AA.
XX
AC AD019051;
XX
DT 26-AUG-2004 (first entry)
XX
DE Murine antibody M1 Fab fragment heavy chain.
XX
KW Mouse; antibody; M1 Fab; heavy chain; Fab; scFv; abused drug; morphine;
KW THC; amphetamine; environmental hazard; toxic compound;
KW microbial process; metabolic process; drug monitoring;
KW pharmacological research.
XX
OS Mus sp.
XX
PN WO2004046733-A1.
XX
PD 03-JUN-2004.
XX
PF 17-NOV-2003; 2003WO-FI000875.
XX
PR 18-NOV-2002; 2002FI-00002048.
XX
PA (VALM ) VALTION TEKNIILLINEN TUTKIMUSKESKUS.
XX
PI Pulli T, Hoeyhtyae M, Takkinen K, Soederlund H;
XX
PS MPI; 2004-420710/39.

```

```

XX
PT Non-competitive immunoassay for small analyte, useful for assaying drug
PT of abuse (e.g., morphine), comprises reacting a sample of analyte with a
PT reagent pair comprising a first binding partner and a second binding
PT partner.
XX
PS Claim 21; SEQ ID NO 2; 35pp; English.
XX
CC The invention relates to a non-competitive immunoassay for a small
CC analyte, comprising reacting a sample containing the analyte with a
CC reagent pair comprising a first binding partner that binds to the analyte
CC and a second binding partner that binds to the complex of the analyte and
CC the first binding partner, and determining the binding of the second
CC binding partner, thus indicating the presence of the analyte in the
CC sample. The first and second binding partners are antibody fragments Fab
CC or scFv. The reagent pair is useful in a non-competitive immunoassay for
CC a small analyte, particularly for assaying drugs of abuse e.g., morphine,
CC THC or amphetamine. The immunoassay is useful for detecting environmental
CC hazards, toxic compounds in food and feed, chemicals indicative of
CC ongoing processes (e.g., microbial processes in buildings, metabolic
CC processes of living organisms) and in clinical tests, drug monitoring and
CC pharmacological research. This sequence represents the murine antibody M1
CC Fab fragment heavy chain, used in the method of the invention.
XX
SQ Sequence 236 AA;

Query Match          55.3%; Score 62.5; DB 8; Length 236;
Best Local Similarity 34.5%; Pred. No. 19;
Matches 10; Conservative 12; Mismatches 4; Indels 3; Gaps 1;

QY 2 SHPOFE---KXXXXXXXXXXWSHPOFEK 27
   :|||
DB 208 AHPASTKVDKXIVRDCGTSMSHPOFEK 236

RESULT 13
AD019053
ID AD019053 standard; protein; 236 AA.
XX
AC AD019053;
XX
DT 26-AUG-2004 (first entry)
XX
DE Murine antibody M2 Fab fragment heavy chain.
XX
KW Mouse; antibody; M2 Fab; heavy chain; Fab; scFv; abused drug; morphine;
KW THC; amphetamine; environmental hazard; toxic compound;
KW microbial process; metabolic process; drug monitoring;
KW pharmacological research.
XX
OS Mus sp.
XX
PN WO2004046733-A1.
XX
PD 03-JUN-2004.
XX
PF 17-NOV-2003; 2003WO-FI000875.
XX
PR 18-NOV-2002; 2002FI-00002048.
XX
PA (VALM ) VALTION TEKNIILLINEN TUTKIMUSKESKUS.
XX
PI Pulli T, Hoeyhtyae M, Takkinen K, Soederlund H;
XX
PS MPI; 2004-420710/39.
XX
PT Non-competitive immunoassay for small analyte, useful for assaying drug
PT of abuse (e.g., morphine), comprises reacting a sample of analyte with a
PT reagent pair comprising a first binding partner and a second binding
PT partner.
XX
PS Claim 21; SEQ ID NO 4; 35pp; English.

```





*This Page Blank (uspto)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 03.9634 Seconds  
(without alignments)  
105.489 Million cell updates/sec

Title: SEQ15  
Perfect score: 113  
Sequence: 1 wshpgefxxxxxxwshpgefek 27

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 328044528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing files 45 summaries

Database : Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep:\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep:\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep:\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep:\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep:\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep:\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09B\_PUBCOMB.pep:\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09C\_PUBCOMB.pep:\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_NEW\_PUB.pep:\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10B\_PUBCOMB.pep:\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep:\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep:\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10\_NEW\_PUB.pep:\*  
18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep:\*  
19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep:\*  
20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	98.5	87.2	24	US-10-026-578B-3	Sequence 3, Appl1
2	98.5	87.2	36	US-10-026-578B-10	Sequence 10, Appl1
3	85.5	75.7	36	US-10-026-578B-11	Sequence 11, Appl1
4	69.5	61.5	24	US-09-809-517A-31	Sequence 31, Appl1
5	69.5	61.5	24	US-10-026-578B-4	Sequence 4, Appl1
6	69.5	61.5	25	US-09-809-517A-34	Sequence 34, Appl1
7	69.5	61.5	245	US-10-887-228A-1	Sequence 1, Appl1
8	69.5	61.5	246	US-10-887-228A-9	Sequence 9, Appl1
9	69.5	61.5	252	US-10-887-228A-5	Sequence 5, Appl1
10	68	60.2	134	US-10-437-963-104065	Sequence 104065,
11	68	60.2	158	US-10-424-599-184769	Sequence 184769,
12	68	60.2	212	US-10-424-599-184075	Sequence 184075,
13	68	60.2	252	US-10-437-963-200150	Sequence 200150,

14	67.5	59.7	646	US-10-628-432-49	Sequence 49, Appl1
15	67.5	59.7	858	US-10-628-432-37	Sequence 27, Appl1
16	66	58.4	485	US-10-628-432-47	Sequence 47, Appl1
17	66	58.4	697	US-10-628-432-24	Sequence 24, Appl1
18	64	56.6	21	US-09-809-517A-30	Sequence 30, Appl1
19	64	56.6	22	US-09-809-517A-33	Sequence 33, Appl1
20	62	54.9	117	US-09-977-137A-4	Sequence 4, Appl1
21	62	54.9	117	US-09-977-137A-5	Sequence 5, Appl1
22	62	54.9	117	US-09-977-137A-7	Sequence 7, Appl1
23	62	54.9	117	US-09-977-137A-8	Sequence 8, Appl1
24	62	54.9	117	US-09-977-137A-9	Sequence 9, Appl1
25	62	54.9	117	US-09-977-137A-10	Sequence 10, Appl1
26	62	54.9	117	US-09-977-137A-11	Sequence 11, Appl1
27	62	54.9	117	US-09-977-137A-12	Sequence 12, Appl1
28	62	54.9	118	US-09-977-137A-6	Sequence 6, Appl1
29	62	54.9	633	US-10-628-432-53	Sequence 53, Appl1
30	62	54.9	661	US-10-358-983-8	Sequence 8, Appl1
31	62	54.9	763	US-10-358-283-15	Sequence 15, Appl1
32	62	54.9	845	US-10-628-432-40	Sequence 40, Appl1
33	59.5	52.7	56	US-09-833-245-1103	Sequence 1103, Ap
34	59.5	52.7	56	US-09-833-245-1105	Sequence 1105, Ap
35	59	52.2	108	US-10-425-114-37298	Sequence 37298, A
36	58	51.3	792	US-10-437-963-180912	Sequence 180912,
37	57.5	50.9	699	US-10-408-765A-434	Sequence 434, App
38	57.5	50.9	699	US-10-741-600-851	Sequence 851, App
39	57.5	50.9	699	US-10-741-600-852	Sequence 852, App
40	57	50.4	36	US-10-424-599-146965	Sequence 146965,
41	57	50.4	385	US-10-437-963-115643	Sequence 115643,
42	57	50.4	472	US-09-369-735B-8	Sequence 8, Appl1
43	56	49.6	100	US-10-424-599-258116	Sequence 258116,
44	56	49.6	529	US-10-437-963-132697	Sequence 132697,
45	56	49.6	706	US-10-437-963-132699	Sequence 132699,

## ALIGNMENTS

RESULT 1  
US-10-026-578B-3  
Sequence 3, Application US/10026578B  
Publication No. US20030083474A1  
GENERAL INFORMATION:  
APPLICANT: IBA (GmbH)  
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta  
FILE REFERENCE: 100810.01US1  
CURRENT APPLICATION NUMBER: US/10/026,578B  
CURRENT FILING DATE: 2002-11-11  
PRIOR APPLICATION NUMBER: DE 101 13 776.1  
PRIOR FILING DATE: 2001-03-21  
PRIOR APPLICATION NUMBER: PCT/EP01/11846  
PRIOR FILING DATE: 2001-10-12  
NUMBER OF SEQ ID NOS: 14  
SOFTWARE: PatentIn version 3.1  
SEQ ID NO 3  
LENGTH: 24  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Synthetic Peptide  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (9)..(9)  
OTHER INFORMATION: X represents a single amino acid at the position indicated  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (10)..(10)  
OTHER INFORMATION: X represents a single amino acid at the position indicated  
FEATURE:  
NAME/KEY: MISC FEATURE  
LOCATION: (11)..(11)  
OTHER INFORMATION: X represents a single amino acid at the position indicated  
FEATURE:

```

? NAME/KEY: MISC_FEATURE
? LOCATION: (12)-(12)
? OTHER INFORMATION: X represents a single amino acid at the position indicated
? FEATURE:
? NAME/KEY: MISC_FEATURE
? LOCATION: (13)-(13)
? OTHER INFORMATION: X represents a single amino acid at the position indicated
? FEATURE:
? NAME/KEY: MISC_FEATURE
? LOCATION: (14)-(14)
? OTHER INFORMATION: X represents a single amino acid at the position indicated
? FEATURE:
? NAME/KEY: MISC_FEATURE
? LOCATION: (15)-(15)
? OTHER INFORMATION: X represents a single amino acid at the position indicated
? FEATURE:
? NAME/KEY: MISC_FEATURE
? LOCATION: (16)-(16)
? OTHER INFORMATION: X represents a single amino acid at the position indicated
? FEATURE:
? NAME/KEY: misc_feature
? OTHER INFORMATION: Artificial sequence represents peptide binding module
? S-10-026-578B-3

```

Query Match	87.2%	Score 98.5;	DB 14,	Length 24;
Best Local Similarity	88.9%;	Pred. No. 4e-05;		
Matches 24; Conservative	0;	Mismatches	3;	Gaps 1;

QY	Db
1 WSHPQFEKXXXXXXXXXXXXWSHPQFEK 27	1 WSHPQFEK-----XXXXXXXXXXXXWSHPQFEK 24

```

      RESULT 2
      US-10-026-578B-10
      ; Sequence 10, Application US/10026578B
      ; Publication No. US20030083474A1
      ; GENERAL INFORMATION:
      ; APPLICANT: IBA (GmbH)
      ; APPLICANT: Schmidt, Thomas
      ; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Tags
      ; FILE REFERENCE: 100810.01US1
      ; CURRENT APPLICATION NUMBER: US/10/026.578B
      ; CURRENT FILING DATE: 2002-11-11
      ; PRIOR APPLICATION NUMBER: DE 101 13 776.1
      ; PRIOR FILING DATE: 2001-03-21
      ; PRIOR APPLICATION NUMBER: PCT/EP01/11846
      ; PRIOR FILING DATE: 2001-10-12
      ; NUMBER OF SEQ ID NOS: 14
      ; SOFTWARE: Patentin version 3.1
      ; SEQ ID NO 10
      ; LENGTH: 36
      ; TYPE: PRT
      ; ORGANISM: Artificial Sequence
      ; FEATURE:
      ; OTHER INFORMATION: Synthetic Peptide
      ; FEATURE:
      ; NAME/KEY: MISC FEATURE
      ; LOCATION: (9)..(28)
      ; OTHER INFORMATION: x represents a single amino acid at each of the positions indicated
      ; OTHER INFORMATION: ed: some of the amino acids may be missing. Where amino acids are
      ; OTHER INFORMATION: missing, the total numbers of x will be no less than 5
      ; FEATURE:
      ; NAME/KEY: misc_feature
      ; OTHER INFORMATION: Artificial sequence represents peptide binding module
      ; US-10-026-578B-10

```

Query Match	87.2%	Score 98.5	DB 14	Length 36
Best Local Similarity	75.0%	Pred. No. 6.8e-05		
Matches 27, Conservative	0	Mismatches 0	Indels 9	Gaps 1
OY	1	MSHPOFEK-----XXXXXXXXXXXXMSHPOFEK	27	

Db 1 WSHPQFEKXXXXXXXXXXXXXXXXXXXXXXXXXXXXXWSHPQFEK 36

```

RESULT 3
US-10-026-578B-11
; Sequence 11, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Tr
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; CURRENT FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 11
; LENGTH: 36
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; FEATURE:
; NAME/KEY: MISC_FEATURE
; LOCATION: (9)..(28)
; OTHER INFORMATION: Of the five (GGGS) repeats between position 9 and 28, one or more
; OTHER INFORMATION: repeats may be missing. However, at least one (GGGS) repeat will
; OTHER INFORMATION: be present
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-11

```

Query Match	75.7%;	Score 85.5;	DB 14;	Length 36;
Best Local Similarity	41.7%;	Pred. No. 0.0032;		
Matches 15;	Conservative 11;	Mismatches 1;	Indels 9;	Gaps 1;

```
Qy      1 WSHDQFEK-----XXXXXXXXXXXXWSHPQFEK 27
        |||||
        :~::~:~:::|||||
Db      1 WSHDQFEKGSGGSGGGSGGGSGGGSGGSQSHDQFEK 36
```

```

RESULT 4
US-09-809-517A-31
; Sequence 31, Application US/09809517A
; Patent No. US20020034733A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: Co-US20020034733A1e1 methods for displaying (poly)peptides/prote
; TITLE OF INVENTION: Particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; CURRENT FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 31
; LENGTH: 24
; TYPE: PRT
; ORGANSIM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial sequence: synthetic module
; US-09-809-517A-31

```

Query Match	61.5%;	Score 69.5;	DB 9;	Length 24;
Best Local Similarity	45.8%;	Pred. NO. 0.22;		

```

Matches 11; Conservative 12; Mismatches 0; Indels 1; Gaps 1;

QY 5 QFE-KXXXXXXXXXXWSHPQPEK 27
   :|||:|||||:|||||:|||||:
   :|||:|||||:|||||:|||||:
DB 1 EFEQKLISEDLNGAPWMSHPQPEK 24

RESULT 5
US-10-026-578B-4
; Sequence 4, Application US/10026578B
; Publication No. US20030083474A1
; GENERAL INFORMATION:
; APPLICANT: IBA (GmbH)
; APPLICANT: Schmidt, Thomas
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
; FILE REFERENCE: 100810.01US1
; CURRENT APPLICATION NUMBER: US/10/026,578B
; PRIOR FILING DATE: 2002-11-11
; PRIOR APPLICATION NUMBER: DE 101 13 776.1
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: PCT/EP01/11846
; PRIOR FILING DATE: 2001-10-12
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: Patentin version 3.1
; SEQ ID NO 4
; LENGTH: 24
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Synthetic Peptide
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (9)..(9)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; NAME/KEY: MISC FEATURE
; LOCATION: (10)..(10)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (11)..(11)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (12)..(12)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (13)..(13)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (14)..(14)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (15)..(15)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (16)..(16)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (17)..(17)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (18)..(18)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; NAME/KEY: MISC FEATURE
; LOCATION: (22)..(22)

OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (23)..(23)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: MISC FEATURE
; LOCATION: (24)..(24)
; OTHER INFORMATION: X represents a single amino acid at the position indicated
; FEATURE:
; NAME/KEY: misc feature
; OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-4

Query Match 61.5%; Score 69.5; DB 14; Length 24;
Best Local Similarity 87.5%; Pred. No. 0.22;
Matches 21; Conservative 0; Mismatches 0; Indels 3; Gaps 1;

QY 1 WSHPQEKXXXXXXXXXXWSHPQ 24
   :|||:|||||:|||||:|||||:
   :|||:|||||:|||||:|||||:
DB 1 WSHPQEKXXXXXXXXXX--HPQ 21

RESULT 6
US-09-809-517A-34
; Sequence 34, Application US/09809517A
; Patent No. US20020034733A1
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. US20020034733A1 methods for displaying (poly)peptides/prote
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34

Query Match 61.5%; Score 69.5; DB 9; Length 25;
Best Local Similarity 45.8%; Pred. No. 0.23;
Matches 11; Conservative 12; Mismatches 0; Indels 1; Gaps 1;

QY 5 QFE-KXXXXXXXXXXWSHPQPEK 27
   :|||:|||||:|||||:|||||:
   :|||:|||||:|||||:|||||:
DB 2 EFEQKLISEDLNGAPWMSHPQPEK 25

RESULT 7
US-10-887-228A-1
; Sequence 1, Application US/10887228A
; Publication No. US20050037402A1
; GENERAL INFORMATION:
; APPLICANT: Scherfing AG
; TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio
; FILE REFERENCE: S30569US
; CURRENT APPLICATION NUMBER: US/10/887,228A
; PRIOR FILING DATE: 2004-07-09
; PRIOR APPLICATION NUMBER: DE 103 31 093.2
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US 60/478,262
; PRIOR FILING DATE: 2003-07-16
; NUMBER OF SEQ ID NOS: 12

```

```
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 1
/ LENGTH: 245
/ TYPE: PRT
/ ORGANISM: artificial
/ FEATURE:
/ OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-1

Query Match      61.5%; Score 69.5; DB 17; Length 245;
Best Local Similarity 45.8%; Pred. No. 4.4;
Matches 11; Conservative 12; Mismatches 0; Indels 1; Gaps 1;

Cy 5 OFE-KXXXXXXSHPOPEK 27
Db 222 EFEOKLISEDLNGAPWHPQPEK 245

RESULT 8
US-10-887-228A-9
/ Sequence 9, Application US/10887228A
/ Publication No. US20050037402A1
/ GENERAL INFORMATION:
/ APPLICANT: Schering AG
/ TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio
/ FILE REFERENCE: S30569US
/ CURRENT APPLICATION NUMBER: US/10/887,228A
/ PRIOR FILING DATE: 2004-07-09
/ PRIOR APPLICATION NUMBER: DE 103 31 093.2
/ PRIOR FILING DATE: 2003-07-09
/ PRIOR APPLICATION NUMBER: US 60/478,262
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 9
/ LENGTH: 246
/ TYPE: PRT
/ ORGANISM: artificial
/ FEATURE:
/ OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-9

Query Match      61.5%; Score 69.5; DB 17; Length 246;
Best Local Similarity 45.8%; Pred. No. 4.4;
Matches 11; Conservative 12; Mismatches 0; Indels 1; Gaps 1;

Cy 5 OFE-KXXXXXXSHPOPEK 27
Db 223 EFEOKLISEDLNGAPWHPQPEK 246

RESULT 9
US-10-887-228A-5
/ Sequence 5, Application US/10887228A
/ Publication No. US20050037402A1
/ GENERAL INFORMATION:
/ APPLICANT: Schering AG
/ TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio
/ FILE REFERENCE: S30569US
/ CURRENT APPLICATION NUMBER: US/10/887,228A
/ PRIOR FILING DATE: 2004-07-09
/ PRIOR APPLICATION NUMBER: DE 103 31 093.2
/ PRIOR FILING DATE: 2003-07-09
/ PRIOR APPLICATION NUMBER: US 60/478,262
/ NUMBER OF SEQ ID NOS: 12
/ SOFTWARE: PatentIn version 3.1
/ SEQ ID NO 5
/ LENGTH: 252
/ TYPE: PRT
/ ORGANISM: artificial
```

```
/ FEATURE:
/ OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-5

Query Match      61.5%; Score 69.5; DB 17; Length 252;
Best Local Similarity 45.8%; Pred. No. 4.5;
Matches 11; Conservative 12; Mismatches 0; Indels 1; Gaps 1;

Cy 5 OFE-KXXXXXXSHPOPEK 27
Db 229 EFEOKLISEDLNGAPWHPQPEK 252

RESULT 10
US-10-437-963-104065
/ Sequence 104065, Application US/10437963
/ Publication No. US20040123343A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J.
/ APPLICANT: Kovalic, David K.
/ APPLICANT: Zhou, Yihua
/ APPLICANT: Cao, Yongwei
/ APPLICANT: Wu, Wei
/ APPLICANT: Bouharov, Andrey A.
/ APPLICANT: Barbazuk, Brad
/ APPLICANT: Li, Ping
/ TITLE OF INVENTION: Rice Nucleic Acid Molecules and Other Molecules Associated With
/ FILE REFERENCE: 38-21(53221)B
/ CURRENT APPLICATION NUMBER: US/10/437,963
/ CURRENT FILING DATE: 2003-05-14
/ NUMBER OF SEQ ID NOS: 204966
/ SEQ ID NO 104065
/ LENGTH: 134
/ TYPE: PRT
/ ORGANISM: Oryza sativa
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT4530_101434C.1.pap
US-10-437-963-104065

Query Match      60.2%; Score 68; DB 16; Length 134;
Best Local Similarity 37.5%; Pred. No. 3.1;
Matches 9; Conservative 13; Mismatches 2; Indels 0; Gaps 0;

Cy 3 HPOYEKXXXXXXSHPOPE 26
Db 57 HPOYEKKIGGIDYLTGALPEFE 80

RESULT 11
US-10-424-599-184769
/ Sequence 184769, Application US/10424599
/ Publication No. US20040031072A1
/ GENERAL INFORMATION:
/ APPLICANT: La Rosa, Thomas J
/ APPLICANT: Kovalic, David K
/ APPLICANT: Zhou Yihua
/ APPLICANT: Cao Yongwei
/ TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With
/ FILE REFERENCE: 38-21(53223)B
/ CURRENT APPLICATION NUMBER: US/10/424,599
/ CURRENT FILING DATE: 2003-04-28
/ NUMBER OF SEQ ID NOS: 285684
/ SEQ ID NO 184769
/ LENGTH: 158
/ TYPE: PRT
/ ORGANISM: Glycine max
/ FEATURE:
/ OTHER INFORMATION: Clone ID: PAT_MRT3847_137862C.1.pap
US-10-424-599-184769

Query Match      60.2%; Score 68; DB 15; Length 158;
```



This Page Blank (uspto)

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 30.7317 Seconds  
(without alignments) 65.585 Million cell updates/sec

Title: SEQ15

Perfect score: 113  
Sequence: 1 whnpqfexkxxxxxxxxxxwhnpqfex 27

Scoring table: BLOSUM62GX  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued Patents AA:\*

1: /cgn2\_6/prodata/1/1aa/5B\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/1aa/6B\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	69.5	61.5	24	US-09-809-517A-31	Sequence 31, Appl
2	69.5	61.5	25	US-09-809-517A-34	Sequence 34, Appl
3	64	56.6	21	US-09-809-517A-30	Sequence 30, Appl
4	64	56.6	22	US-09-809-517A-33	Sequence 33, Appl
5	62	54.9	117	US-09-977-137A-4	Sequence 4, Appl
6	62	54.9	117	US-09-977-137A-5	Sequence 5, Appl
7	62	54.9	117	US-09-977-137A-7	Sequence 7, Appl
8	62	54.9	117	US-09-977-137A-8	Sequence 8, Appl
9	62	54.9	117	US-09-977-137A-9	Sequence 9, Appl
10	62	54.9	117	US-09-977-137A-10	Sequence 10, Appl
11	62	54.9	117	US-09-977-137A-11	Sequence 11, Appl
12	62	54.9	117	US-09-977-137A-12	Sequence 12, Appl
13	62	54.9	118	US-09-977-137A-6	Sequence 6, Appl
14	61	54.0	434	US-09-252-991A-25931	Sequence 25931, A
15	60	53.1	386	US-08-895-707-7	Sequence 7, Appl
16	57.5	50.9	699	US-09-538-092-995	Sequence 995, App
17	56	49.6	125	US-09-621-976-4303	Sequence 4303, Ap
18	56	49.6	306	US-09-248-796A-16127	Sequence 16127, A
19	56	49.6	2547	US-09-058-489-35	Sequence 35, Appl
20	56	49.6	2547	US-09-538-092-1374	Sequence 1374, Ap
21	55.5	49.1	647	US-09-248-796A-15400	Sequence 15400, A
22	55	48.7	84	US-09-270-767-36127	Sequence 36127, A
23	55	48.7	84	US-09-270-767-36127	Sequence 36127, A
24	54	47.8	223	US-09-489-039A-12356	Sequence 12356, A
25	54	47.8	394	US-09-248-796A-17857	Sequence 17857, A
26	53	46.9	10	US-09-809-517A-6	Sequence 6, Appl
27	53	46.9	400	US-07-989-991A-2	Sequence 2, Appl

28	53	46.9	400	3	US-08-621-255-2	Sequence 2, Appl
29	53	46.9	400	3	US-09-352-574-2	Sequence 2, Appl
30	53	46.9	638	4	US-09-252-991A-25205	Sequence 25205, A
31	53	46.9	979	1	US-08-308-881-6	Sequence 6, Appl
32	53	46.9	979	2	US-09-058-263-6	Sequence 6, Appl
33	53	46.9	979	2	US-09-059-099-6	Sequence 6, Appl
34	53	46.9	979	3	US-09-058-264-6	Sequence 6, Appl
35	53	46.9	979	4	US-09-455-962-6	Sequence 6, Appl
36	53	46.9	979	5	PCT-US95-06530-6	Sequence 6, Appl
37	52.5	46.5	389	2	US-08-811-949-65	Sequence 65, Appl
38	52	46.0	27	2	US-08-310-912A-47	Sequence 47, Appl
39	52	46.0	27	3	US-08-841-089-47	Sequence 47, Appl
40	52	46.0	27	3	US-09-301-085-47	Sequence 47, Appl
41	52	46.0	27	5	PCT-US95-04570-47	Sequence 47, Appl
42	52	46.0	27	5	PCT-US95-04589-47	Sequence 47, Appl
43	52	46.0	249	4	US-09-270-767-48926	Sequence 48926, A
44	52	46.0	250	4	US-09-270-767-33709	Sequence 33709, A
45	52	46.0	659	4	US-09-252-991A-26013	Sequence 26013, A

#### ALIGNMENTS

```
RESULT 1
US-09-809-517A-31
Sequence 31, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on t
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
SOFTWARE: PatentIn version 3.0
SEQ ID NO 31
LENGTH: 24
TYPE: PRT
ORGANISM: artificial sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-31
Query Match      61.5%  Score 69.5;  DB 4;  Length 24;
Best Local Similarity 45.8%;  Pred. No. 0.0088;
Matches 11;  Conservative 12;  Mismatches 0;  Indels 1;  Gaps 1;
OY 5 OPE-KXXXXXXWXXSHPOPEK 27
:|||||:
Db 1 EPEQKISBEDNGAPWSHPOPEK 24
:|||||:
RESULT 2
US-09-809-517A-34
Sequence 34, Application US/09809517A
Patent No. 6753136
GENERAL INFORMATION:
APPLICANT: Lohning, Corinna
TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
FILE REFERENCE: MORPHO/11
CURRENT APPLICATION NUMBER: US/09/809,517A
PRIOR FILING DATE: 2001-03-15
PRIOR APPLICATION NUMBER: EP 99114072.4
PRIOR FILING DATE: 1999-07-20
PRIOR APPLICATION NUMBER: EP 00103551.8
PRIOR FILING DATE: 2000-02-18
NUMBER OF SEQ ID NOS: 41
```

```
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34

Query Match          61.5%; Score 69.5; DB 4; Length 25;
Best Local Similarity 45.8%; Pred. No. 0.0094;
Matches 11; Conservative 12; Mismatches 0; Indels 1; Gaps 1;

QY 5 EFKXXXXXXXXXXWSHPQFEK 27
Db 2 EFDYKDDDDKGAPMWSHPQFEK 25

RESULT 3
US-09-809-517A-30.
; Sequence 30, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 30
; LENGTH: 21
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30

Query Match          56.6%; Score 64; DB 4; Length 21;
Best Local Similarity 42.9%; Pred. No. 0.045;
Matches 9; Conservative 11; Mismatches 1; Indels 0; Gaps 0;

QY 7 EKKXXXXXXXXXXWSHPQFEK 27
Db 1 EFDYKDDDDKGAPMWSHPQFEK 21

RESULT 4
US-09-809-517A-33
; Sequence 33, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on b
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patentin version 3.0
; SEQ ID NO 33
; LENGTH: 22
; TYPE: PRT
; ORGANISM: artificial sequence
```

```
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33

Query Match          56.6%; Score 64; DB 4; Length 22;
Best Local Similarity 42.9%; Pred. No. 0.049;
Matches 9; Conservative 11; Mismatches 1; Indels 0; Gaps 0;

QY 7 EKKXXXXXXXXXXWSHPQFEK 27
Db 2 EFDYKDDDDKGAPMWSHPQFEK 22

RESULT 5
US-09-977-137A-4
; Sequence 4, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: Methods
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 4
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4

Query Match          54.9%; Score 62; DB 4; Length 117;
Best Local Similarity 42.1%; Pred. No. 1.2;
Matches 8; Conservative 11; Mismatches 0; Indels 0; Gaps 0;

QY 9 XXXXXXXXXXXXXWSHPQFEK 27
Db 99 ARKGNVSCPSAMSHQFEK 117

RESULT 6
US-09-977-137A-5
; Sequence 5, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: Methods
; CURRENT APPLICATION NUMBER: US/09/977,137A
; PRIOR FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 5
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-5

Query Match          54.9%; Score 62; DB 4; Length 117;
Best Local Similarity 42.1%; Pred. No. 1.2;
Matches 8; Conservative 11; Mismatches 0; Indels 0; Gaps 0;
```





```

; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 11
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-11

```

```

Query Match      54.9%; Score 62; DB 4; Length 117;
Best Local Similarity 42.1%; Pred. No. 1.2;
Matches      8; Conservative 11; Mismatches      0; Indels      0; Gaps      0;

```

```

QY      9 XXXXXXXXXXXXWHPQFEK 27
        :XXXXXXXXXXXXXXXXXXXX
Db      99 ARKGNVCPGSAWHPQFEK 117

```

```

RESULT 12
US-09-977-137A-12
; Sequence 12, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 12
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-12

```

```

Query Match      54.9%; Score 62; DB 4; Length 117;
Best Local Similarity 42.1%; Pred. No. 1.2;
Matches      8; Conservative 11; Mismatches      0; Indels      0; Gaps      0;

```

```

QY      9 XXXXXXXXXXXXWHPQFEK 27
        :XXXXXXXXXXXXXXXXXXXX
Db      99 ARKGNVCPGSAWHPQFEK 117

```

```

RESULT 13
US-09-977-137A-6
; Sequence 6, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; APPLICANT: Cagliati, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18

```

```

; SOFTWARE: Patentin Ver. 2.0
; SEQ ID NO 6
; LENGTH: 118
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-6

```

```

Query Match      54.9%; Score 62; DB 4; Length 118;
Best Local Similarity 42.1%; Pred. No. 1.2;
Matches      8; Conservative 11; Mismatches      0; Indels      0; Gaps      0;

```

```

QY      9 XXXXXXXXXXXXWHPQFEK 27
        :XXXXXXXXXXXXXXXXXXXX
Db      100 ARKGNVCPGSAWHPQFEK 118

```

```

RESULT 14
US-09-252-991A-25931
; Sequence 25931, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 25931
; LENGTH: 434
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-25931

```

```

Query Match      54.0%; Score 61; DB 4; Length 434;
Best Local Similarity 28.0%; Pred. No. 12;
Matches      7; Conservative 15; Mismatches      3; Indels      0; Gaps      0;

```

```

QY      1 WSHPQFEKXXXXXXXXXXXXWHPQF 25
        :XXXXXXXXXXXXXXXXXXXX
Db      100 WSIYPAFERAMRHGVARDGYLPYAF 124

```

```

RESULT 15
US-08-895-707-7
; Sequence 7, Application US/08895707
; Patent No. 6077700
; GENERAL INFORMATION:
; APPLICANT: (Pharmacia & Upjohn, Co.)
; APPLICANT: alternatively, for U.S. filing:
; APPLICANT: Hollingsworth, Robert A.
; APPLICANT: Sharma, Satish K.
; APPLICANT: Rank, Kenneth B.
; APPLICANT: Evans, David B.
; TITLE OF INVENTION: Special Constructs and Complexes of
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pharmacia & Upjohn Company
; STREET: 301 Henrietta Street
; CITY: Kalamazoo
; STATE: MI
; COUNTRY: USA
; ZIP: 49001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: floppy disk
; COMPUTER: IBM PC compatible

```

OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Rel. #1.0, Ver. #1.25/WordPerfect 5.2+  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/895,707  
FILING DATE:  
CLASSIFICATION: 435  
ATTORNEY/AGENT INFORMATION:  
NAME: Mooton, Thomas A.  
REGISTRATION NUMBER: 35,004  
REFERENCE/DOCKET NUMBER: 6054  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 616-833-7914  
TELEFAX: 616-833-8897  
INFORMATION FOR SEQ ID NO: 7:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 386 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: peptide  
HYPOTHETICAL: NO  
ANTI-SENSE: NO  
FRAGMENT TYPE: N-terminal  
US-08-895-707-7

Query Match 53.1%; Score 60; DB 3; Length 386;  
Best Local Similarity 136.4%; Pred. No. 14;  
Matches 8; Conservative 11; Mismatches 3; Indels 0; Gaps 0;  
QY 4 POFKXXXXXXXXXXWSHPQF 25  
DB 363 POSGKKQSGPEMASAKRHQF 384

Search completed: March 2, 2005, 12:25:39  
Job time: 31.7317 secs

*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 21.3984 Seconds  
(without alignments)  
125.901 Million cell updates/sec

Title: SEQ16  
Perfect score: 114  
Sequence: 1 wshpgefekxxxxxxxxxxwshpgefek 28

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 79:\*  
1: Df1:\*  
2: Pf1:\*  
3: Pf3:\*  
4: Pf4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	60	52.6	435	2 A44308	Antho-Rfamid prec
2	60	52.6	1107	2 T25450	hypothetical prote
3	59.5	52.2	601	2 T11677	probable transcrip
4	58	50.9	313	2 T40926	conserved hypothet
5	58	50.9	652	2 T34497	hypothetical prote
6	57	50.0	151	2 T00816	hypothetical prote
7	57	50.0	206	2 S76147	hypothetical prote
8	56.5	49.6	342	2 A45462	C 4.2.3.9 aristolo
9	56	49.1	103	2 E90973	hypothetical prote
10	56	49.1	103	2 A85821	unknown protein en
11	56	49.1	103	2 C85713	unknown protein en
12	56	49.1	103	2 A90904	hypothetical prote
13	56	49.1	334	2 A39172	Antho-Rfamid neur
14	56	49.1	335	2 AD3492	acetate-CoA 1
15	56	49.1	583	2 S30014	hypothetical prote
16	56	49.1	699	2 A54660	histidine rich cal
17	56	49.1	732	2 H83376	1,4-alpha-glucan b
18	56	49.1	2408	2 T24483	hypothetical prote
19	55.5	48.7	213	2 AD2533	hypothetical prote
20	55	48.2	234	2 T30473	late expression fa
21	55	48.2	253	2 AG2559	transposase al1807
22	55	48.2	398	2 AB2622	succinyl-diaminop
23	55	48.2	398	2 B97404	succinyl-diaminop
24	55	48.2	527	2 C88042	protein F56D12.4
25	55	48.2	624	2 T26148	hypothetical prote
26	55	48.2	995	2 T39724	probable pre-mRNA
27	55	48.2	3071	2 T45584	hypothetical prote
28	54	47.4	317	2 S49005	non-structural pro
29	54	47.4	323	2 B70960	probable nadr proc

30	54	47.4	484	1 SYRZET	glutamate-tRNA lig
31	54	47.4	507	2 T24944	hypothetical prote
32	54	47.4	537	2 YRMSB6	tyrosinase-related
33	54	47.4	617	2 F75484	hypothetical prote
34	54	47.4	805	2 E82297	c-di-GMP phosphodi
35	54	47.4	801	2 H83737	glucosidase BH0704
36	53.5	46.9	211	2 B71058	hypothetical prote
37	53.5	46.9	553	1 S29861	hybrid cluster 14F
38	53.5	46.9	1091	2 S01998	contactin precuro
39	53	46.5	139	2 T41693	conserved hypothet
40	53	46.5	228	2 S73041	hypothetical prote
41	53	46.5	230	2 B83879	hypothetical prote
42	53	46.5	246	2 C83694	hypothetical prote
43	53	46.5	282	2 T27554	hypothetical prote
44	53	46.5	290	2 T43351	nuclear receptor N
45	53	46.5	297	2 A45442	transport vesicle

## ALIGNMENTS

## RESULT 1

A44308  
Antho-Rfamid precursor - sea anemone (Anthopleura elegantissima)  
C:Species: Anthopleura elegantissima  
C>Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 09-Jul-2004  
C:Accession: A44308  
R:Schmützler, C.; Darner, D.; Diekhoff, D.; Grimmlikhuijzen, C.J.  
J. Biol. Chem. 267, 22534-22541, 1992  
A>Title: Identification of a novel type of processing sites in the precursor for the sea  
A:Reference number: A44308; MUID:93054550; PMID:1429603  
A:Accession: A44308  
A:Status: preliminary  
A:Molecule type: mRNA; protein  
A:Residues: 1-435 <SCH>  
A:Cross-References: UNIPROT:P10419; GB:M96269; NID:G155702; PIDN:AA27738.1; PID:G155703  
A>Note: sequence extracted from NCBI backbone (NCBIN:117102, NCBIF:117104)  
C:Keywords: neuropeptide

## Query Match

Best Local Similarity 52.6%; Score 60; DB 2; Length 435;  
Matches 10; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

QY 1 WSHPOPEKXXXXXXXXXXWSHPPEK 28

DB 66 FSDPQFWKGRFSDPQFWKGRFSDPQFWK 93

## RESULT 2

T25450  
hypothetical protein B0412.3 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004  
C:Accession: T25450

R:Bentley, D.  
submitted to the EMBL Data Library, December 1996  
A:Description: The sequence of C. elegans cosmid B0412.  
A:Reference number: Z20037

A:Accession: T25450  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA

A:Residues: 1-1107 <BEN>  
A:Cross-References: UNIPROT:P09084; EMBL:U80953; PIDN:AA852556.1; GSPDB:GN00021; CESP:B0

A:Experimental source: strain Bristol N2; clone B0412  
C:Genetics:  
A:Gene: CESP:B0412.3  
A:Map position: 3

A:Introns: 39/2, 70/1, 133/2, 288/3, 386/1, 470/3, 568/3, 631/2, 733/2, 819/3, 931/1, 96  
C:Superfamily: Caenorhabditis elegans hypothetical protein B0412.3

## Query Match

Best Local Similarity 52.6%; Score 60; DB 2; Length 1107;  
Matches 6; Conservative 15; Mismatches 7; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXXWSHPQEK 28  
 DB 76 WFOKLDKVPALIVFIDLEWDPMSWP 103

## RESULT 3

probable transcription factor - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C>Date: 16-Jul-1999 #sequence\_revision 16-Jul-1999 #text\_change 09-Jul-2004  
 R:Seeger, K.; Harris, D.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
 A:Submitted to the EMBL Data Library, September 1998  
 A:Reference number: Z17313  
 A:Accession: T11677  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-601 <SEB>  
 A:Cross-references: UNIPROT:O74345; EMBL:AL031536; NID:el319499; PID:el319504  
 C:Genetic: experimental source: strain 972h(-)  
 A:Map position: IIR  
 A:Introns: 39/1; 74/3; 106/2  
 A:Note: SPBC21D10.05C

Query Match 52.2%; Score 59.5; DB 2; Length 601;  
 Best Local Similarity 31.0%; Pred. No. 35;  
 Matches 9; Conservative 13; Mismatches 2; Indels 5; Gaps 1;

QY 1 WSHPOFEKXXXXXXXXXXWSHP 24  
 DB 68 WSNQLEKMKMGWINANRYMNPPLSP 96

## RESULT 4

conserved hypothetical protein SPCC1281.07c - fission yeast (Schizosaccharomyces pombe)  
 C:Species: Schizosaccharomyces pombe  
 C>Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 09-Jul-2004  
 C:Accession: T40926  
 R:Volckaert, G.; Wood, V.; Rajandream, M.A.; Barrell, B.G.  
 A:Submitted to the EMBL Data Library, January 1999  
 A:Reference number: Z21957  
 A:Accession: T40926  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-313 <VOL>  
 A:Cross-references: UNIPROT:O94524; EMBL:AL035218; PIDN:CAA22828.1; GSPDB:GN00068; SPDB:  
 C:Genetic: experimental source: strain 972h-; cosmid c1281  
 A:Gene: SPDB:SPCC1281.07C  
 A:Map position: 3  
 C:Superfamily: glutathione S-transferase

Query Match 50.9%; Score 58; DB 2; Length 313;  
 Best Local Similarity 28.0%; Pred. No. 26;  
 Matches 7; Conservative 13; Mismatches 5; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXXWSHPQ 25  
 DB 271 WGRPAFHETTDPKHKKHYTQSHQ 295

## RESULT 5

hypothetical protein ZK1248.15 - Caenorhabditis elegans  
 C:Species: Caenorhabditis elegans  
 C>Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 09-Jul-2004  
 C:Accession: T34497  
 R:Latreille, P.  
 A:Submitted to the EMBL Data Library, June 1995  
 A:Description: The sequence of C. elegans cosmid ZK1248.

A:Reference number: Z21534

A:Accession: T34497

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-652 <LAT>

A:Cross-references: UNIPROT:Q23425; EMBL:U29244; PIDN:AACT1096.1; GSPDB:GN00020; CESP:ZL

A:Experimental source: strain Bristol N2; clone ZK1248

C:Genetic: CESP:ZK1248.15

A:Map position: 2

A:Introns: 29/2; 107/3; 172/2; 264/3; 310/3; 360/1; 443/1; 575/3

C:Superfamily: Caenorhabditis elegans hypothetical protein ZK1248.15; WW repeat homology

F:198-236/Domain: WW repeat homology <WWR>

Query Match 50.9%; Score 58; DB 2; Length 652;  
 Best Local Similarity 21.4%; Pred. No. 61;  
 Matches 6; Conservative 15; Mismatches 7; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXXWSHPQEK 28  
 DB 207 WWHTEKKKKFYNDKXESLMDHNTTRK 234

## RESULT 6

hypothetical protein At2g41600 [imported] - Arabidopsis thaliana  
 N:Alternate names: hypothetical protein T3266.12  
 C:Species: Arabidopsis thaliana (mouse-ear cress)  
 C>Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004  
 C:Accession: T00816; G84843  
 R:Rounsley, S.D.; Lin, X.; Ketchum, K.A.; Crosby, M.L.; Brandon, R.C.; Sykes, S.M.; Kaul  
 A:Submitted to the EMBL Data Library, November 1997  
 A:Description: Arabidopsis thaliana chromosome II BAC T3266 genomic sequence.  
 A:Reference number: Z14163  
 A:Accession: T00816  
 A:Status: translated from GB/EMBL/DBJ  
 A:Molecule type: DNA  
 A:Residues: 1-151 <ROU>

A:Cross-references: UNIPROT:O22222; EMBL:AC002510; NID:G2618683; PID:G2618695  
 A:Experimental source: cultivar Columbia

R:Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.L.; Town, C.D.; Fujii, C.Y.;

M.; Koo, H.; Moffat, K.S.; Cronlin, L.A.; Shen, M.; Venakten, S.E.; Umayam, L.; Taiton, L.

enus, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Frazer, C.M.; Venter, J

Nature 402, 761-768, 1999

A:Title: Sequence and analysis of chromosome 2 of the plant Arabidopsis thaliana.

A:Reference number: A84420; WUID:20083487; PMID:10617197

A:Accession: G84843

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-151 <STO>

A:Cross-references: GB:AE002093; NID:G2618695; PIDN:AAB84342.1; GSPDB:GN00139

C:Genetic:

A:Gene: T3266.12; At2g41600

A:Map position: 2

A:Introns: 38/3

C:Superfamily: Arabidopsis thaliana hypothetical protein At2g41600

Query Match 50.0%; Score 57; DB 2; Length 151;  
 Best Local Similarity 25.0%; Pred. No. 15;  
 Matches 6; Conservative 15; Mismatches 3; Indels 0; Gaps 0;

QY 2 SHPOFEKXXXXXXXXXXWSHPQ 25  
 DB 33 SHRFQGVETGSLGPKLMDSPR 56

## RESULT 7

hypothetical protein - Synechocystis sp. (strain PCC 6803)  
 C:Species: Synechocystis sp.  
 C:Accession: PCC 6803  
 C>Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
 C:Accession: S76147

R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; O. K.; Okumura, S.; Shimizu, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A>Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
S.  
A:Reference number: 874322; MUID:97061201; PMID:8905231  
A:Accession: S76147  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-206 <KAN>  
A:Cross-references: UNIPROT:E74312; EMBL:D90914; GB:AB001339; NID:g1653477; PIDN:BA01840  
A>Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

Query Match 50.0%; Score 57; DB 2; Length 206;  
Best Local Similarity 25.0%; Pred. No. 22;  
Matches 6; Conservative 16; Mismatches 2; Indels 0; Gaps 0;

Qy 2 SHPOEKXXXXXXXXXXSHPO 25  
|||:|||||:|||||:|||||:  
Db 58 SHPKVQELLELSQWSEGVWSSPE 81

RESULT 8  
A:Accession: A45462  
C 4.2.3.9 aristochoene synthase ( ) - Penicillium roqueforti  
C:Species: Penicillium roqueforti  
C>Date: 21-Sep-1993 #sequence\_revision 25-Apr-1997 #text\_change 09-Jul-2004  
C:Accession: A45462  
R:Proctor, R.H.; Hohn, T.M.  
J. Biol. Chem. 268, 4543-4548, 1993  
A>Title: Aristochoene synthase, isolation, characterization, and bacterial expression  
A:Reference number: A45462; MUID:93179472; PMID:8440737  
A:Accession: A45462  
A:Molecule type: DNA  
A:Residues: 1-342 <PRO>  
A:Cross-references: UNIPROT:O03471; GB:L05193; NID:g169177; PIDN:AA33694.1; PID:g169178  
A>Note: sequence extracted from NCBI backbone (NCBI:125997)  
C:Comment: This enzyme is a sesquiterpene cyclase involved in synthesis of the sesquiterpene  
C:Genetics:  
A:Gene: At11  
C:Superfamily: aristochoene synthase  
C:Keywords: transferase

Query Match 49.6%; Score 56.5; DB 2; Length 342;  
Best Local Similarity 23.3%; Pred. No. 47;  
Matches 7; Conservative 16; Mismatches 4; Indels 3; Gaps 1;

Qy 1 WS---HPOEKXXXXXXXXXXSHPO 27  
|||:|||||:|||||:|||||:  
Db 46 WSYLCHPRVKVQDEVDCGFLENMKRPFK 75

RESULT 9  
E90973  
hypothetical protein E90973 [imported] - Escherichia coli (strain O157:H7, substrain RI  
C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
C:Accession: E90973  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.  
Sasawara, N.; Yasunaga, T.; Kuhara, S.; Shibata, T.; Hattori, M.; Shinagawa, H.  
DNA Res. 8, 11-22, 2001  
A>Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene  
A:Reference number: A99629; MUID:21156231; PMID:11258796  
A:Accession: E90973  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-103 <HAY>  
A:Cross-references: UNIPROT:Q8X4V0; GB:BA000007; PIDN:BA036180.1; PID:g13362225; GSPDB:C  
A:Experimental source: strain O157:H7, substrain RIMD 0509952  
C:Genetics:  
A:Gene: EC82757

Query Match 49.1%; Score 56; DB 2; Length 103;

Best Local Similarity 22.7%; Pred. No. 14;  
Matches 5; Conservative 17; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSHPOEKXXXXXXXXXXWS 22  
|||:|||||:|||||:|||||:  
Db 17 WTHPEYKFCDDREYISTEEFN 38

RESULT 10  
A5821  
unknown protein encoded within prophage CP-9330 [imported] - Escherichia coli (strain O1  
C:Species: Escherichia coli  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: A5821  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lm, A.; Diallanata, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: A85821  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-103 <STO>  
A:Cross-references: UNIPROT:Q8X4V0; GB:AE005174; NID:g12516136; PIDN:AA657029.1; GSPDB:G  
A:Experimental source: strain O157:H7, substrain ED933  
C:Genetics:  
A:Gene: Z3120

Query Match 49.1%; Score 56; DB 2; Length 103;  
Best Local Similarity 22.7%; Pred. No. 14;  
Matches 5; Conservative 17; Mismatches 0; Indels 0; Gaps 0;

Qy 1 WSHPOEKXXXXXXXXXXWS 22  
|||:|||||:|||||:|||||:  
Db 17 WTHPEYKFCDDREYISTEEFN 38

RESULT 11  
C85713  
unknown protein encoded within prophage CP-9330 [imported] - Escherichia coli (strain O1  
C:Species: Escherichia coli  
C>Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004  
C:Accession: C85713  
R:Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew  
Miller, L.; Grobeck, E.J.; Davis, N.W.; Lm, A.; Diallanata, E.; Potamousis, K.; Apodaca,  
Nature 409, 529-533, 2001  
A>Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.  
A:Reference number: A85480; MUID:21074935; PMID:11206551  
A:Accession: C85713  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-103 <STO>  
A:Cross-references: UNIPROT:Q8X549; GB:AE005174; NID:g12515046; PIDN:AA656167.1; GSPDB:C  
A:Experimental source: strain O157:H7, substrain ED933  
C:Genetics:  
A:Gene: Z2097

Query Match 49.1%; Score 56; DB 2; Length 103;  
Best Local Similarity 27.3%; Pred. No. 14;  
Matches 6; Conservative 15; Mismatches 1; Indels 0; Gaps 0;

Qy 7 ECKXXXXXXXXXXSHPOEK 28  
|||:|||||:|||||:|||||:  
Db 3 EIKEMPVARDGXYWTHPEYK 24

RESULT 12  
A90904  
hypothetical protein E90904 [imported] - Escherichia coli (strain O157:H7, substrain RI  
C:Species: Escherichia coli  
C>Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004  
C:Accession: A90904  
R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.

gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.  
 DNA Res. 8, 11-22, 2001  
 A>Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gend  
 A:Reference number: A59629; MUID:21156231; PMID:11258796  
 A:Accession: A5904  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-103 <HAY>  
 A:Cross-references: UNIPROT:O8X549; GB:BA000007; PIDN:BA35624.1; PID:913361667; GSPDB:G  
 A:Experimental source: strain O157:H7, substrain R1MD 050952  
 C:Genetics:  
 A:Gene: ECs2201

Query Match 49.1%; Score 56; DB 2; Length 103;  
 Best Local Similarity 27.3%; Pred. No. 14;  
 Matches 6; Conservative 15; Mismatches 1; Indels 0; Gaps 0;  
 Oy 7 EKXXXXXXXXXXXXXSHPOPEK 28  
 Db 3 EIKEMPVVRDGYGWTPEYER 24

RESULT 13  
 A39172  
 Antho-Ramide neuropeptide 19 repeat precursor - sea anemone (*Calliactis parasitica*)  
 C:Species: *Calliactis parasitica*  
 C>Date: 07-Feb-1992 #sequence\_revision 07-Feb-1992 #text\_change 09-Jul-2004  
 C:Accession: A39172  
 R:Darmer, D.; Schmutzler, C.; Diekhoff, D.; Grimmelshuizen, C.J.P.  
 Proc. Natl. Acad. Sci. U.S.A. 88, 2555-2559, 1991  
 A>Title: Primary structure of the precursor for the sea anemone neuropeptide Antho-RFam  
 A:Reference number: A39172; MUID:91172845; PMID:1706527  
 A:Accession: A39172  
 A:Status: Preliminary  
 A:Molecule type: mRNA  
 A:Residues: 1-334 <DAR>  
 A:Cross-references: UNIPROT:Q01133; GB:M59166; NID:9156133; PIDN:AAA27878.1; PID:9156134  
 C:Keywords: neuropeptide

Query Match 49.1%; Score 56; DB 2; Length 334;  
 Best Local Similarity 36.0%; Pred. No. 53;  
 Matches 9; Conservative 13; Mismatches 3; Indels 0; Gaps 0;  
 Oy 4 POPEXXXXXXXXXXXXXSHPOPEK 28  
 Db 56 PQFWKGRFSDPQFWKGRFSDPQFWK 80

RESULT 14  
 AD3492  
 acetate-CoA ligase (EC 6.2.1.16) [imported] - *Brucella melitensis* (strain 16M)  
 C:Species: *Brucella melitensis*  
 C>Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 09-Jul-2004  
 C:Accession: AD3492  
 R:DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mijer, C.; Los, T.; Ivanova,  
 Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002  
 A>Title: The genome sequence of the facultative intracellular pathogen *Brucella melitensis*  
 A:Reference number: AD3252; PMID:1175688  
 A:Accession: AD3492  
 A:Status: Preliminary  
 A:Molecule type: DNA  
 A:Residues: 1-335 <KUR>  
 A:Cross-references: UNIPROT:O8YF6; GB:AE00917; PIDN:AAL53103.1; PID:917983968; GSPDB:G  
 A:Experimental source: strain 16M  
 C:Genetics:  
 A:Gene: BME11922  
 A:Map position: I  
 C:Keywords: acid-thiol ligase; coenzyme A

Query Match 49.1%; Score 56; DB 2; Length 335;  
 Best Local Similarity 26.9%; Pred. No. 53;  
 Matches 7; Conservative 13; Mismatches 6; Indels 0; Gaps 0;

Oy 1 WSHPOPEXXXXXXXXXXXXXSHPOPEK 26  
 Db 164 WMDAGEKIQAYFPERFDWCHGDF 189

RESULT 15  
 S30014  
 hypothetical protein YKL012w - yeast (*Saccharomyces cerevisiae*)  
 N:Alternate names: hypothetical protein YKL165  
 C:Species: *Saccharomyces cerevisiae*  
 C>Date: 07-May-1993 #sequence\_revision 07-May-1993 #text\_change 09-Jul-2004  
 C:Accession: S30014; S37823; S37825  
 R:Pascolo, S.; Ghazvini, M.; Boyer, J.; Collea, L.; Thierry, A.; Dujon, B.  
 Yeast 8, 987-995, 1992  
 A>Title: The sequence of a 9.3 kb segment located on the left arm of the yeast chromosome  
 A:Reference number: S30013; MUID:93127732; PMID:1481574  
 A:Accession: S30014  
 A:Molecule type: DNA  
 A:Residues: 1-583 <PAS>  
 A:Cross-references: UNIPROT:P33203; GB:S53418; NID:9263497; PIDN:AMB24902.1; PID:9263498  
 R:Boyer, J.; Pascolo, S.; Richard, G.F.; Ghazvini, M.; Collea, L.; Thierry, A.; Monne  
 submitted to the Protein Sequence Database, March 1994  
 A:Reference number: S37813  
 A:Accession: S37823  
 A:Molecule type: DNA  
 A:Residues: 1-583 <BOY>  
 A:Cross-references: EMBL:Z28012; NID:9485996; PIDN:CAA81847.1; PID:9485997; MIPS:YKL012w  
 A:Experimental source: strain S288C  
 R:Wiemann, S.; Voss, H.; Schwager, C.; Rupp, T.; Grothues, D.; Sengen, C.; Stegemann, J.  
 submitted to the Protein Sequence Database, March 1994  
 A:Reference number: S37825  
 A:Accession: S37825  
 A:Molecule type: DNA  
 A:Residues: 1-583 <WIE>  
 A:Cross-references: EMBL:Z28012; NID:9485996; PIDN:CAA81847.1; PID:9485997; MIPS:YKL012w  
 A:Experimental source: strain S288C  
 C:Genetics:  
 A:Gene: SGD:PPP40  
 A:Cross-references: SGD:S0001495; MIPS:YKL012w  
 A:Map position: 11L  
 F:1-35/Domain: WW repeat homology #status atypical <WW1>  
 F:39-76/Domain: WW repeat homology <WW2>

Query Match 49.1%; Score 56; DB 2; Length 583;  
 Best Local Similarity 32.0%; Pred. No. 16+02;  
 Matches 8; Conservative 13; Mismatches 4; Indels 0; Gaps 0;  
 Oy 1 WSHPOPEXXXXXXXXXXXXXSHPOPEK 25  
 Db 67 WTPAEKKEVEPIAEQKHDVSHAQ 91

Search completed: March 2, 2005, 12:29:02  
 Job time: 22.3984 secs



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 104.033 Seconds  
(without alignments)  
137.824 Million cell updates/sec

Title: SEQ16  
Perfect score: 114  
Sequence: 1 wshpqlfexxxxxxxxxxwshpqlfex 28

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 1612378 seqs, 512079187 residues  
Total number of hits satisfying chosen parameters: 1612378

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt 03: \*  
1: uniprot\_sprot: \*  
2: uniprot\_trembl: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	67	58.8	1749	2	Q8T0W6	Q8T0W6 echinococcus
2	64	56.1	704	2	Q8CJD2	Q8CJD2 rattus norvegicus
3	63	55.3	743	2	Q8CH90	Q8CH90 rattus norvegicus
4	61	53.5	304	2	Q74J66	Q74J66 lactobacillus
5	61	53.5	869	2	Q62MT8	Q62MT8 burkholderia
6	61	53.5	869	2	Q6JYS0	Q6JYS0 burkholderia
7	60.5	53.1	107	2	Q6Z315	Q6Z315 oryza sativa
8	60.5	53.1	719	2	Q87NM4	Q87NM4 oryza sativa
9	60	52.6	429	1	FMK2_ANTTEL	FMK2_ANTTEL
10	60	52.6	435	1	FMK1_ANTTEL	FMK1_ANTTEL
11	60	52.6	506	2	Q73TFS	Q73TFS mycobacterium
12	60	52.6	508	1	MATK_MARSC	MATK_MARSC
13	60	52.6	508	2	Q9BBG0	Q9BBG0 marichthys
14	60	52.6	508	2	Q9BBG3	Q9BBG3 vanoroyenella
15	60	52.6	650	2	Q83DG9	Q83DG9 oeserya coulteri
16	60	52.6	671	2	Q6YQD6	Q6YQD6 coxiella burnetii
17	60	52.6	1015	2	Q6C664	Q6C664 onion yellows
18	60	52.6	1107	2	Q6C664	Q6C664 yarrowia lipolytica
19	59.5	52.2	601	2	P90984	P90984 caenorhabditis
20	59	51.8	759	2	Q8RYC2	Q8RYC2 schizosaccharomyces
21	59	51.8	2333	2	Q8RSP8	Q8RSP8 oryza sativa
22	58	50.9	166	2	Q8RYE4	Q8RYE4 candida glabrata
23	58	50.9	196	2	Q6GCH6	Q6GCH6 staphylococcus
24	58	50.9	297	2	Q891J0	Q891J0 staphylococcus
25	58	50.9	313	2	Q94524	Q94524 bradyrhizobium
26	58	50.9	337	2	Q6VY41	Q6VY41 schizosaccharomyces
27	58	50.9	364	2	Q6FTJ1	Q6FTJ1 vibrio parahaemolyticus
28	58	50.9	431	2	Q7N8A2	Q7N8A2 candida glabrata
29	58	50.0	471	2	Q23425	Q23425 photorhabdus
30	57	50.0	59	2	Q76K22	Q76K22 caenorhabditis
31	57	50.0	151	2	Q22222	Q22222 neurospora crassa

32	57	50.0	192	2	Q6UNF7 oryza sativ
33	57	50.0	206	2	P74312 synechocyst
34	57	50.0	247	2	Q82S44 nitrosomon
35	57	50.0	305	2	Q8W403 oryza sativ
36	57	50.0	305	2	Q6Z0Y9 oryza sativ
37	57	50.0	355	2	Q9VX10 dirosophila
38	57	50.0	541	2	Q987P4 rhizobium l
39	57	50.0	590	2	Q6CVV1 kluyveromyc
40	57	50.0	652	2	Q988Y6 kluyveromyc
41	57	50.0	719	1	GLB2_XANAC
42	57	50.0	1698	1	CUL7_HUMAN
43	56.5	49.6	144	2	Q8ZG44 streptomyce
44	56.5	49.6	224	2	Q8RVV8 coffea arab
45	56.5	49.6	322	2	Q6CSW9 kluyveromyc

## ALIGNMENTS

ID	Q8T0W6	PRELIMINARY;	PRT;	1749 AA.
AC	Q8T0W6;			
DT	01-JUN-2002 (TREMBLrel. 21, Created)			
DT	01-JUN-2002 (TREMBLrel. 21, Last sequence update)			
DT	01-MAR-2004 (TREMBLrel. 26, Last annotation update)			
DE	Putative insulin receptor precursor.			
GN	Name=Irh;			
OS	Echinococcus multilocularis.			
OC	Eukaryota; Metazoa; Platyhelminthes; Cestoda; Eucestoda;			
OC	Cyclophyllidae; Taeniidae; Echinococcus.			
OX	NCBI_Taxid=6211;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RA	MEDLINE=22557731; PubMed=12670515; DOI=10.1016/S0020-7519(02)00265-5;			
RA	Konrad C., Kroner A., Spiliotis M., Zavala-Gongora R., Brehm K.;			
RT	"Identification and molecular characterisation of a gene encoding a			
RT	member of the insulin receptor family in Echinococcus			
RT	multilocularis."			
RL	Int. J. Parasitol. 33:301-312(2003).			
DR	EMBL; AJ458426; CAD30260.1; -.			
DR	HSSP; P08069; 1JOH.			
DR	GO; GO:0016020; C:membrane; IEA.			
DR	GO; GO:0005524; F:ATP binding; IEA.			
DR	GO; GO:0005006; F:epidermal growth factor receptor activity; IEA.			
DR	GO; GO:0004872; F:receptor activity; IEA.			
DR	GO; GO:0016740; F:transferase activity; IEA.			
DR	GO; GO:0006468; P:protein amino acid phosphorylation; IEA.			
DR	GO; GO:0007169; P:transmembrane receptor protein tyrosine kin. . .; IEA.			
DR	InterPro; IPR00494; EGFR_L.			
DR	InterPro; IPR003961; FN_III.			
DR	InterPro; IPR008957; FN_III-like.			
DR	InterPro; IPR006211; Furin-like.			
DR	InterPro; IPR006212; Furin repeat.			
DR	InterPro; IPR009030; Grow_Fac_recept.			
DR	InterPro; IPR000585; Hemopexin.			
DR	InterPro; IPR01009; Kinase-like.			
DR	InterPro; IPR000719; Prot_Kinase.			
DR	InterPro; IPR001245; Tyr_Kinase.			
DR	InterPro; IPR008266; Tyr_Kinase_AS.			
DR	Pfam; PF00757; Furin-like; 1.			
DR	Pfam; PF01030; Recep_L_domain; 1.			
DR	PRINTS; PR00109; TYRKINASE.			
DR	SMART; SMO0060; FN3; 3.			
DR	SMART; SMO0261; FU; 1.			
DR	SMART; SMO0219; TyKc; 1.			
DR	PROSITE; PS00024; HEMOPEXIN; UNKNOWN_1.			
DR	PROSITE; PS00011; PROTEIN_KINASE_DOM; 1.			
DR	PROSITE; PS00109; PROTEIN_KINASE_TYR; 1.			
KW	Kinase; Receptor; Signal; Transferase; Tyrosine-protein kinase.			
FT	SIGNAL 1 32 Potential.			
FT	CHAIN 33 1749 Potential.			
SQ	SEQUENCE 1749 AA; 192538 MW; 9A8308BFE9597DA CRC64;			

Query Match 58.8%; Score 67; DB 2; Length 1749;  
 Best Local Similarity 32.0%; Pred. No. 53;  
 Matches 8; Conservative 15; Mismatches 2; Indels 0; Gaps 0;

## RESULT 2

ID Q8CJD2 PRELIMINARY; PRT; 704 AA.

AC Q8CJD2; 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DE 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 OS Guanylyl cyclase alpha 1 subunit.  
 OC Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nakamura I., Suzuki N.;  
 RL Submitted (NOV-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB096020; BAC24016.1; -.  
 DR HSSP; P30803; IAZS.  
 DR GO; GO:0004383; F:guanylate cyclase activity; IEA.  
 DR GO; GO:0016829; P:lyase activity; IEA.  
 DR GO; GO:0007242; P:intracellular signaling cascade; IEA.  
 DR InterPro; IPR01054; G\_cyclase.  
 DR SMART; SM00211; Guanylate\_cyc; 1.  
 DR PROSITE; PS00452; GUANYLATE\_CYCLASES\_2; 1.  
 DR PROSITE; PS0125; GUANYLATE\_CYCLASES\_2; 1.  
 KW Lyase.  
 SQ SEQUENCE 704 AA; 79188 MW; F73FC97B685456A CRC64;

Query Match 56.1%; Score 64; DB 2; Length 704;  
 Best Local Similarity 39.1%; Pred. No. 48;  
 Matches 9; Conservative 12; Mismatches 2; Indels 0; Gaps 0;

Qy 6 FEKXXXXXXXXXXXXSHPOFEK 28  
 Db 682 FLGKASGVLDLPRGSMWHPQFEK 704

## RESULT 3

ID Q8CH90 PRELIMINARY; PRT; 743 AA.

AC Q8CH90; 01-MAR-2003 (TREMBlrel. 23, Created)  
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)  
 DE 01-MAR-2004 (TREMBlrel. 26, Last annotation update)  
 OS Soluble guanylyl cyclase alpha 2 subunit E219G mutant.  
 OC Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxId=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Nakamura I., Yao Y., Suzuki N.;  
 RL Submitted (DEC-2002) to the EMBL/GenBank/DBJ databases.  
 DR EMBL; AB097860; BAC44887.1; -.  
 DR HSSP; P30803; IAZS.  
 DR GO; GO:0004383; F:guanylate cyclase activity; IEA.  
 DR GO; GO:0016829; F:lyase activity; IEA.  
 DR InterPro; IPR01054; G\_cyclase.  
 DR InterPro; IPR009080; tRNA\_syn\_ia\_bind.  
 DR Pfam; PF00211; Guanylate\_cyc; 1.

DR SMART; SM00044; CYC; 1.  
 DR PROSITE; PS00452; GUANYLATE\_CYCLASES\_2; 1.  
 DR PROSITE; PS0125; GUANYLATE\_CYCLASES\_2; 1.  
 KW Lyase.  
 SQ SEQUENCE 743 AA; 83251 MW; ACF5CS3E0982813A CRC64;

Query Match 55.3%; Score 63; DB 2; Length 743;  
 Best Local Similarity 40.0%; Pred. No. 70;  
 Matches 8; Conservative 12; Mismatches 0; Indels 0; Gaps 0;

Qy 9 XXXXXXXXXXXXXXXSHPOFEK 28  
 Db 724 FLRETSILVRGSMWHPQFEK 743

## RESULT 4

ID Q74JB6 PRELIMINARY; PRT; 304 AA.

AC Q74JB6; 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DE 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 OS Hypothetical protein.  
 GN Ordered locus names=J01161;  
 OC Lactobacillus johnsonii.  
 OC Bacteria; Firmicutes; Lactobacillales; Lactobacillaceae;  
 OC Lactobacillus.  
 OX NCBI\_TaxId=33959;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=NC 533;  
 RC PubMed=14966310;  
 RA PubMed=14966310;  
 RA Fridmore R.D., Berger B., Desiere F., Vilanova D., Barretto C.,  
 RA Pitter A.-C., Zwaalen M.-C., Rouvet M., Altemann E., Barrangou R.,  
 RA Mollet B., Mercenier A., Kleenhammer T., Arigoni F., Schell M.A.;  
 RT "The genome sequence of the probiotic intestinal bacterium  
 RT Lactobacillus johnsonii NC 533".  
 RL Proc. Natl. Acad. Sci. U.S.A. 101:2512-2517(2004).  
 DR EMBL; AB017203; AAS08983.1; -.  
 DR GO; GO:0001760; F:aminocarboxymuconate-semialdehyde decarboxy. .; IEA.  
 DR GO; GO:0008152; P:metabolism; IEA.  
 DR InterPro; IPR002160; Prot inh kunz-19.  
 DR PROSITE; PS00283; SOYBEAN\_KUNITZ; UNKNOWN\_1.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 304 AA; 34432 MW; A81B9DC23410B5A4 CRC64;

Query Match 53.5%; Score 61; DB 2; Length 304;  
 Best Local Similarity 29.2%; Pred. No. 46;  
 Matches 7; Conservative 14; Mismatches 3; Indels 0; Gaps 0;

Qy 4 POFXXXXXXXXXXXXSHPOFE 27  
 Db 140 FBEFPIILKAAEHLPLWLPVFD 163

## RESULT 5

ID Q62MT8 PRELIMINARY; PRT; 869 AA.

AC Q62MT8; 25-OCT-2004 (TREMBlrel. 28, Created)  
 DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)  
 DE 25-OCT-2004 (TREMBlrel. 28, Last annotation update)  
 OS DNA topoisomerase III (EC 5.99.1.2).  
 GN Name=topB; ORFNames=BMA0139.  
 OS Burkholderia mallei ATCC 23344.  
 OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;  
 OC Burkholderiaceae; Burkholderia.  
 OX NCBI\_TaxId=243160;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA STRAIN=ATCC 23344;  
 RA Nierman W.C., Deshazer D., Kim H.S., Tetteijn H., Nelson K.E.,  
 RA Feldblyum T., Ulrich R.L., Ronning C.M., Brinkac L.M., Daugherty S.C.,

```

RA Davidson T.D., Debroy R.T., Dimitrov G., Dodson R.J., Durkin A.S.,
RA Gwin M.L., Haft D.H., Khouri H., Koloney J.F., Madupu R.,
RA Mohammad Y., Nelson W.C., Radune D., Romero C.M., Satria S.,
RA Selengut U., Shandlin C., Sullivan S.A., White O., Yu Y., Zatar N.,
RA Zhou L., Fraser C.M.;
RT "Structural flexibility in the Burkholderia mallei genome.";
RL Proc. Natl. Acad. Sci. U.S.A. 101:14247-14251 (2004).
DR EMBL; CP000010; AAU48686.1; -.
KM Isomerase.
SQ SEQUENCE 869 AA; 96949 MW; 32EB90771AA5E556 CRC64;

Query Match
Best Local Similarity 53.5%; Score 61; DB 2; Length 869;
Matches 8; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXXWSHPQFE 27
DB 237 WDPKFKKDEFPKRDRLMLPAAE 263

RESULT 6
Q63YS0 PRELIMINARY; PRT; 869 AA.
AC Q63YS0;
DT 25-OCT-2004 (TREMBlrel. 28, Created)
DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)
DE Putative DNA topoisomerase III (EC 5.99.1.2).
GN ORFNames=BPSL0118;
OS Burkholderia pseudomallei K96243.
OC Bacteria; Proteobacteria; Betaproteobacteria; Burkholderiales;
OC Burkholderiaceae; Burkholderia.
OK NCBI_TaxId=272560;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=K96243;
RX PubMed=15377794;
RA Holden M.T.G., Tilball R.W., Peacock S.J., Cerdano-Tarraga A.M.,
RA Atkins T., Crossman L.C., Pitt T., Churcher C., Mungall K.,
RA Bentley S.D., Sebatina M., Thomson N.R., Baeson N., Beacham I.R.,
RA Brooks K., Brown K.A., Brown N.F., Challis G.L., Cherevach I.,
RA Chillingworth T., Cronin A., Crossat B., Davis P., Deshazer D.,
RA Felwell T., Fraser A., Hance Z., Hauser H., Holtroyd S., Jagsels K.,
RA Keith K.E., Maddison M., Moule S., Price C., Quail M.A.,
RA Rabinovitch E., Rutherford K., Sanders M., Simmonds M.,
RA Songelylala S., Stevens K., Tumapa S., Vesaratchaveat M.,
RA Whitehead S., Yeats C., Barrett B.G., Oyston P.C.F., Parkhill J.;
RT "Genomic plasticity of the causative agent of melioidosis,
RT Burkholderia pseudomallei."
RL Proc. Natl. Acad. Sci. U.S.A. 101:14240-14245 (2004).
DR EMBL; BX571965; CAH34104.1; -.
KM Isomerase.
SQ SEQUENCE 869 AA; 96945 MW; 32ED56B71AA5E152 CRC64;

Query Match
Best Local Similarity 53.5%; Score 61; DB 2; Length 869;
Matches 8; Conservative 14; Mismatches 5; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXXXXWSHPQFE 27
DB 237 WDPKFKKDEFPKRDRLMLPAAE 263

RESULT 7
Q62315 PRELIMINARY; PRT; 107 AA.
AC Q62315;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE DCL protein-like.
GN Name=OJ1004 A11.16-2; Synonyms=P0519D10.35-2;
OS Oryza sativa (japonica cultivar-group).

```

```

OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;
OC Eriaristidae; Oryzaeae; Oryza.
OK NCBI_TaxId=39947;
RN (1)
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Katayose Y.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
RM [2]
RP SEQUENCE FROM N.A.
RA Sasaki T., Matsumoto T., Yamamoto K.;
RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP005287; BAD17315.1; -.
DR EMBL; AP004817; BAD17127.1; -.
SQ SEQUENCE 107 AA; 12839 MW; 00F06AF3A12B0EB1 CRC64;

Query Match
Best Local Similarity 53.1%; Score 60.5; DB 2; Length 107;
Matches 9; Conservative 14; Mismatches 1; Indels 1; Gaps 1;

QY 3 HPQFEKXXXXXXXXXXWSHPQFE 27
DB 30 HPQFEKXIGCGIDYLVG-LHPEFE 53

RESULT 8
Q87NW4 PRELIMINARY; PRT; 719 AA.
AC Q87NW4;
DT 01-JUN-2003 (TREMBlrel. 24, Created)
DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)
DT 01-MAR-2004 (TREMBlrel. 26, Last annotation update)
DE Hypothetical protein VP1754.
GN Ordered locus names=VP1754;
OS Vibrio parahaemolyticus.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Vibrionales;
OC Vibrionaceae; Vibrio.
OK NCBI_TaxId=670;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=RIMD 2210633 / Serotype O3:K6;
RX MEDLINE=22508454; PubMed=12620739; DOI=10.1016/S0140-6736(03)12659-1;
RA Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K.,
RA Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S.,
RA Yasunaga T., Honda T., Shinagawa H., Hattori W., Iida T.;
RT "Genome sequence of Vibrio parahaemolyticus: a pathogenic mechanism
RT distinct from that of V. cholerae."
RL Lancet 361:743-749 (2003).
DR EMBL; AP005079; BAC60017.1; -.
DR GO; GO:0016020; C:membrane; IEA.
DR GO; GO:0004871; F:signal transducer activity; IEA.
DR GO; GO:0007165; F:signal transduction; IEA.
DR InterPro; IPR001633; EAL.
DR InterPro; IPR001630; GGDEF.
DR InterPro; IPR003660; His_kin_HAMP.
DR Pfam; PF00563; EAL; 1.
DR Pfam; PF00590; GGDEF; 1.
DR Pfam; PF00672; HAMP; 1.
DR SMART; SM00267; DUF1; 1.
DR SMART; SM00052; DUF2; 1.
DR SMART; SM00304; HAMP; 1.
DR TIGRfams; TIGR00254; GGDEF; 1.
DR PROSITE; PS50883; EAL; 1.
DR PROSITE; PS50887; GGDEF; 1.
DR PROSITE; PS50885; HAMP; 1.
KM Complete proteome.
SQ SEQUENCE 719 AA; 81471 MW; CAB9E68F0C0C6F1 CRC64;

Query Match
Best Local Similarity 53.1%; Score 60.5; DB 2; Length 719;
Matches 7; Conservative 16; Mismatches 2; Indels 1; Gaps 1;

QY 1 WSHPOFEKXXXXXXXXXXWSHPQ 25

```

Db 473 WYQPKEDLKTQYGVGEALVKNHPE 498

RESULT 9

FMRL ANTEL STANDARD; PRT; 429 AA.

AC 016994;

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 05-JUN-2004 (Rel. 44, Last annotation update)

DE Antho-Rfamide neuropeptides type 2 precursor.

OS Anthopleura elegantissima (Sea anemone).

OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actinaria;

OC Nyantheae; Actinidae; Anthopleura.

OX NCBI\_TaxID=6110;

RP SEQUENCE FROM N.A.

RX MEDLINE=93054550; PubMed=1429603;

RA Schmutzler C., Darmer D., Diekhoff D., Grimmelikhuijsen C.J.P.;

RT "Identification of a novel type of processing sites in the precursor

RT from Anthopleura elegantissima." Antho-Rfamide (<Glu-Gly-Arg-Phe-NH2)

RL J. Biol. Chem. 267:22534-22541 (1992).

[2]

RP PARTIAL SEQUENCE (ANTHO-RFAMIDE).

RX MEDLINE=87092339; PubMed=2879288;

RA Grimmelikhuijsen C.J.P., Graff D.;

RT "Isolation of pyroglu-Gly-Arg-Phe-NH2 (Antho-Rfamide), a neuropeptide

RT from sea anemones."

RL Proc. Natl. Acad. Sci. U.S.A. 83:9817-9821 (1986).

CC -1- FUNCTION: Not known but it could act as a transmitter at

CC neuromuscular synapses.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: Belongs to the FARP (FMRFamide related peptide)

CC family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M99170; AAA27738.1; --

DR InterPro: IPR002544; FARP.

DR Pfam: PF01581; FARP; 14.

KW Annotation: Direct protein sequencing; Neuropeptide; Repeat; Signal.

FT SIGNAL 1 22

FT PEPTIDE 234 237 Antho-Rfamide.

FT PEPTIDE 242 245 Antho-Rfamide.

FT PEPTIDE 250 253 Antho-Rfamide.

FT PEPTIDE 258 261 Antho-Rfamide.

FT PEPTIDE 266 269 Antho-Rfamide.

FT PEPTIDE 274 277 Antho-Rfamide.

FT PEPTIDE 290 293 Antho-Rfamide.

FT PEPTIDE 298 301 Antho-Rfamide.

FT PEPTIDE 306 309 Antho-Rfamide.

FT PEPTIDE 322 325 Antho-Rfamide.

FT PEPTIDE 330 333 Antho-Rfamide.

FT PEPTIDE 343 346 Antho-Rfamide.

FT PEPTIDE 356 359 Antho-Rfamide.

FT PEPTIDE 369 372 Antho-Rfamide.

FT MOD\_RES 237 237 Phenylalanine amide (G-238 provides amide

FT MOD\_RES 245 245 group).

FT MOD\_RES 253 253 Phenylalanine amide (G-246 provides amide

FT MOD\_RES 261 261 group).

FT MOD\_RES 269 269 Phenylalanine amide (G-254 provides amide

FT MOD\_RES 269 269 group).

FT MOD\_RES 269 269 Phenylalanine amide (G-262 provides amide

FT MOD\_RES 269 269 group).

FT MOD\_RES 269 269 Phenylalanine amide (G-270 provides amide

FT MOD\_RES 269 269 group).

FT MOD\_RES 277 277 (group).

FT MOD\_RES 293 293 Phenylalanine amide (G-278 provides amide

FT MOD\_RES 301 301 group).

FT MOD\_RES 301 301 Phenylalanine amide (G-294 provides amide

FT MOD\_RES 301 301 group).

FT MOD\_RES 309 309 Phenylalanine amide (G-302 provides amide

FT MOD\_RES 309 309 group).

FT MOD\_RES 325 325 Phenylalanine amide (G-310 provides amide

FT MOD\_RES 325 325 group).

FT MOD\_RES 333 333 Phenylalanine amide (G-326 provides amide

FT MOD\_RES 333 333 group).

FT MOD\_RES 346 346 Phenylalanine amide (G-334 provides amide

FT MOD\_RES 346 346 group).

FT MOD\_RES 359 359 Phenylalanine amide (G-347 provides amide

FT MOD\_RES 359 359 group).

FT MOD\_RES 372 372 Phenylalanine amide (G-360 provides amide

FT MOD\_RES 372 372 group).

FT MOD\_RES 372 372 Phenylalanine amide (G-373 provides amide

FT MOD\_RES 372 372 group).

SO SEQUENCE 429 AA; 50564 MW; 7C54F5C606D537F4 CRC64;

Query Match 52.6%; Score 60; DB 1; Length 429;

Best Local Similarity 35.7%; Pred. No. 96;

Matches 10; Conservative 14; Mismatches 4; Indels 0; Gaps 0;

QY 1 WSHPOFEKXXXXXXXKXWSHPQFK 28

DB 66 FSDPQFMKGRFSDPQFMKGRFSDPQFK 93

FMRL ANTEL STANDARD; PRT; 435 AA.

ID FMRL ANTEL

AC P10419;

DT 01-MAR-1989 (Rel. 10, Created)

DT 01-OCT-1993 (Rel. 27, Last sequence update)

DT 05-JUN-2004 (Rel. 44, Last annotation update)

DE Antho-Rfamide neuropeptides type 1 precursor.

OS Anthopleura elegantissima (Sea anemone).

OC Eukaryota; Metazoa; Cnidaria; Anthozoa; Zoantharia; Actinaria;

OC Nyantheae; Actinidae; Anthopleura.

OX NCBI\_TaxID=6110;

RP SEQUENCE FROM N.A.

RX MEDLINE=93054550; PubMed=1429603;

RA Schmutzler C., Darmer D., Diekhoff D., Grimmelikhuijsen C.J.P.;

RT "Identification of a novel type of processing sites in the precursor

RT from Anthopleura elegantissima." Antho-Rfamide (<Glu-Gly-Arg-Phe-NH2)

RL J. Biol. Chem. 267:22534-22541 (1992).

[2]

RP PARTIAL SEQUENCE (ANTHO-RFAMIDE).

RX MEDLINE=87092339; PubMed=2879288;

RA Grimmelikhuijsen C.J.P., Graff D.;

RT "Isolation of pyroglu-Gly-Arg-Phe-NH2 (Antho-Rfamide), a neuropeptide

RT from sea anemones."

RL Proc. Natl. Acad. Sci. U.S.A. 83:9817-9821 (1986).

CC -1- FUNCTION: Not known but it could act as a transmitter at

CC neuromuscular synapses.

CC -1- SUBCELLULAR LOCATION: Secreted.

CC -1- SIMILARITY: Belongs to the FARP (FMRFamide related peptide)

CC family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL: M99170; AAA27738.1; --

DR PIR: A26666; ECXAA.



```

CC -----
DR EMBL: AB038195; BAB3156.1; -
DR InterPro: IPR008998; Agglutinin.
DR InterPro: IPR000442; Intron_mature2.
DR InterPro: IPR002866; MatK_N.
DR Pfam: PF01348; Intron_mature2; 1.
DR Pfam: PF01824; MatK_N; 1.
DR Chloroplast; mRNA processing.
SQ SEQUENCE 508 AA; 61342 MW; ECCFSB416B0AB5C3 CRC64;

Query Match 52.6%; Score 60; DB 1; Length 508;
Best Local Similarity 28.6%; Pred. No. 1.2e+02;
Matches 8; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

OY 1 WSHPOFEKXXXXXXXXXXSHPOFEK 28
Db 168 WLKDPASLHFLRLEFHEYWNWMSHPSEK 195

RESULT 13
O9BBG0 PRELIMINARY; PRT; 508 AA.
AC O9BBG0;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DE Maturase K.
DE Maturase K.
GN Name=matk;
OS Vanroyenella plumosa.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC eurosid I; Malpighiales; Podostemaceae; Vanroyenella.
OX NCBI_TaxId=51609;
[1]
RP SEQUENCE FROM N.A.
RA Kita Y., Kato M.;
RT "Infrafamilial Phylogeny of the Aquatic Angiosperm Podostemaceae
inferred from the Nucleotide Sequences of the matK Gene.";
RL Plant Biol. 3:156-163(2001).
DR EMBL: AB048378; BAB3398.1; -.
DR GO: GO:0009507; C:chloroplast; IEA.
DR GO: GO:0009507; C:chloroplast; IEA.
DR InterPro: IPR008998; Agglutinin.
DR InterPro: IPR000442; Intron_mature2.
DR InterPro: IPR002866; MatK_N.
DR Pfam: PF01348; Intron_mature2; 1.
DR Pfam: PF01824; MatK_N; 1.
DR Chloroplast.
SQ SEQUENCE 508 AA; 61288 MW; 90229C45138AF4B6 CRC64;

Query Match 52.6%; Score 60; DB 2; Length 508;
Best Local Similarity 28.6%; Pred. No. 1.2e+02;
Matches 8; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

OY 1 WSHPOFEKXXXXXXXXXXSHPOFEK 28
Db 168 WLKDPASLHFLRLEFHEYWNWMSHPSEK 195

RESULT 14
O9BBG3 PRELIMINARY; PRT; 508 AA.
AC O9BBG3;
DT 01-JUN-2001 (TrEMBLrel. 17, Created)
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)
DE Maturase K.
GN Name=matk;
OS Oseya coulteriana.
OC Chloroplast.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;

```

```

OC eurosid I; Malpighiales; Podostemaceae; Oseya.
OX NCBI_TaxId=51602;
[1]
RP SEQUENCE FROM N.A.
RA Kita Y., Kato M.;
RT "Infrafamilial Phylogeny of the Aquatic Angiosperm Podostemaceae
inferred from the Nucleotide Sequences of the matK Gene.";
RL Plant Biol. 3:156-163(2001).
DR EMBL: AB048375; BAB3395.1; -.
DR GO: GO:0009507; C:chloroplast; IEA.
DR GO: GO:0009507; C:chloroplast; IEA.
DR InterPro: IPR008998; Agglutinin.
DR InterPro: IPR000442; Intron_mature2.
DR InterPro: IPR002866; MatK_N.
DR Pfam: PF01348; Intron_mature2; 1.
DR Pfam: PF01824; MatK_N; 1.
DR Chloroplast.
SQ SEQUENCE 508 AA; 61400 MW; 45240C7D31B54CDF CRC64;

Query Match 52.6%; Score 60; DB 2; Length 508;
Best Local Similarity 28.6%; Pred. No. 1.2e+02;
Matches 8; Conservative 12; Mismatches 8; Indels 0; Gaps 0;

OY 1 WSHPOFEKXXXXXXXXXXSHPOFEK 28
Db 168 WLKDPASLHFLRLEFHEYWNWMSHPSEK 195

RESULT 15
O83DG9 PRELIMINARY; PRT; 650 AA.
AC O83DG9;
DT 01-JUN-2003 (TrEMBLrel. 24, Created)
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
DE Acetoacetyl-CoA synthase.
DE Acetoacetyl-CoA synthase.
GN OrderedLocusNames=CBU0766;
OS Coxiella burnetii.
OC Bacteria; Proteobacteria; Gammaproteobacteria; Legionellales;
OC Coxiellaceae; Coxiella.
OX NCBI_TaxId=777;
[1]
RP SEQUENCE FROM N.A.
RA STRAIN=Nine Mile phase I / RSA 493;
RC MERLIN=22608657; PubMed=12704232; DOI=10.1073/pnas.0931379100;
RX Seeharti R., Paulsen I.T., Eisen J.A., Read T.D., Nelson K.E.,
RA Nelson W.C., Ward N.L., Tettein H., Davidsen T.M., Beaman M.J.,
RA DeBoy R.T., Daugherty S.C., Brinkac L.M., Madupu R., Dodson R.J.,
RA Khouri H.M., Lee K.H., Carty H.A., Scanlan D., Heinzen R.A.,
RA Thompson H.A., Samuel J.E., Fraser C.M., Heidelberg J.F.;
RT "Complete genome sequence of the Q-fever pathogen, Coxiella
burnetii.";
RL Proc. Natl. Acad. Sci. U.S.A. 100:5455-5460(2003).
CC -1- SIMILARITY: Belongs to the ATP-dependent AMP-binding enzyme
family.
CC EMBL: AE016962; AAC09302.1; -.
DR TIGR: CB0766; -.
DR GO: GO:0003824; F:catalytic activity; IEA.
DR GO: GO:0016405; F:CoA-ligase activity; IEA.
DR GO: GO:0008299; P:isoprenoid biosynthesis; IEA.
DR GO: GO:0008152; P:metabolism; IEA.
DR InterPro: IPR005914; Acac CoA synth.
DR InterPro: IPR000873; AMP-bind.
DR Pfam: PF00501; AMP-binding; 1.
DR PRINTS: PR00154; AMPBINDING.
DR TIGRFAMs: TIGR01217; ac_ac_CoA_syn; 1.
DR PROSITE: PS00455; AMP_BINDING; 1.
DR Complete proteome.
SQ SEQUENCE 650 AA; 72925 MW; FEDEC2AC651C45A CRC64;

Query Match 52.6%; Score 60; DB 2; Length 650;
Best Local Similarity 21.4%; Pred. No. 1.6e+02;
Matches 6; Conservative 17; Mismatches 5; Indels 0; Gaps 0;

```

```
Oy      1 WSHPOFEKXXXXXXXXXXXXWSHPOFEK 28
         | : | : : : : : : : : | : |
Db      481 WNDPEGRKYYQAYFDKYPNTWAHGDAYK 508
```

Search completed: March 2, 2005, 12:44:32  
Job time : 106.033 secs

*This Page Blank (uspto)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

# OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 126.114 Seconds  
(without alignments)  
85.869 Million cell updates/sec

Title: \$SEQ16  
Perfect score: 114  
Sequence: 1 whnpqfexxxxxxxxxxxxxwhnpqfek 28

Scoring table:  
BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 2105692 seqs, 386760381 residues

Total number of hits satisfying chosen parameters: 2105692

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_16dec04:\*  
1: Geneseq19808:\*  
2: Geneseq19905:\*  
3: Geneseq20008:\*  
4: Geneseq20018:\*  
5: Geneseq20025:\*  
6: Geneseq20038:\*  
7: Geneseq20038:\*  
8: Geneseq20046:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100.5	88.2	35	6	ABP60369
2	100	87.7	36	6	ABP60370
3	98	86.0	24	6	ABP60362
4	69	60.5	24	6	ABP60363
5	68	59.6	763	7	ADB85492
6	67.5	59.2	485	7	AD820249
7	67.5	59.2	697	8	AD820227
8	66	57.9	179	5	AAE24889
9	66	57.9	183	5	AAE24889
10	66	57.9	183	5	AAE24889
11	66	57.9	183	5	AAE24889
12	66	57.9	183	5	AAE24889
13	66	57.9	183	5	AAE24889
14	66	57.9	183	5	AAE24889
15	66	57.9	183	5	AAE24889
16	66	57.9	183	5	AAE24889
17	66	57.9	183	5	AAE24889
18	66	57.9	183	5	AAE24889
19	66	57.9	183	5	AAE24889
20	66	57.9	183	5	AAE24889
21	65	57.0	236	8	AD019051
22	65	57.0	236	8	AD019051
23	65	57.0	254	4	AA846421
24	65	57.0	254	4	AA846421
25	65	57.0	330	8	ADN36400

26	65	57.0	330	8	ADN36408	ADN36408 Human pro
27	65	57.0	396	2	AAW93966	AAW93966 Plasmid p
28	65	57.0	396	4	AAW93966	AAW93966 Plasmid p
29	65	57.0	396	4	AAW93966	AAW93966 Plasmid p
30	65	57.0	400	6	ADA27291	ADA27291 Plasmid p
31	65	57.0	400	6	ADA27292	ADA27292 Plasmid p
32	65	57.0	400	6	ADA00700	ADA00700 Modified
33	65	57.0	400	6	ADA00707	ADA00707 Modified
34	65	57.0	400	6	ADA00701	ADA00701 Modified
35	65	57.0	448	4	AAW97558	AAW97558 Synthetic
36	65	57.0	646	8	AD820251	AD820251 Human agg
37	65	57.0	659	4	AAW94626	AAW94626 Plasmid p
38	65	57.0	659	4	AAW94627	AAW94627 Plasmid p
39	65	57.0	845	8	AD820243	AD820243 Human agg
40	65	57.0	858	8	AD820230	AD820230 Human agg
41	64	56.1	467	8	ADP18461	ADP18461 Arthrobac
42	64	56.1	539	6	ABG74883	ABG74883 Bacteriop
43	64	56.1	539	6	AD139158	AD139158 Enterobac
44	63.5	55.7	633	8	AD820255	AD820255 Human agg
45	63	55.3	117	5	AAW97558	AAW97558 Synthetic

## ALIGNMENTS

RESULT 1	ABP60369	standard, peptide, 35 AA.
ID	ABP60369	standard, peptide, 35 AA.
AC	ABP60369	
XX		
DT	28-MAR-2003	(first entry)
DE	Streptavidin binding peptide SEQ ID NO 10.	
XX		
KW	Streptavidin; protein chip; microtitre plate; detection.	
XX		
OS	Synthetic.	
FH	Key	Location/Qualifiers
FT	Misc-difference 9..27	/label= unknown
FT		/note= "optionally deleted for 1-15 residues"
XX		
PN	DE1013776-A1.	
XX		
PD	02-OCT-2002.	
XX		
PF	21-MAR-2001; 2001DE-01013776.	
PK	21-MAR-2001; 2001DE-01013776.	
XX		
PA	(BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.	
XX		
PI	Schmidt T;	
XX		
DR	WPI; 2003-031166/03.	
XX		
PT	New isolated peptide, useful as affinity purification tag for recombinant protein, comprises at least two high-affinity streptavidin-binding modules.	
XX		
PS	Claim 7, Page 16, 18pp, German.	
XX		
CC	The invention relates to an isolated peptide (I) comprising at least two individual modules separated by 0-50 amino acids, with each containing at least one motif His-Pro-X where X = Gln, Asn or Met. (II), which contain streptavidin binding modules, are useful as affinity handles for purification of recombinant fusion proteins (FP), also for detecting FP, e.g. on protein chips or microtitre plates. The modules in (I) bind strongly to streptavidin, with a co-operative effect that provides stronger binding than a single tag, but are displaced by a competitor. (I) does not interfere with the function of attached proteins (II) (so it	

CC may not be essential to remove (t); facilitates detection and has easily  
CC controllable binding properties. (I) is particularly used for purifying  
CC FP from dilute solution in batch formats (which use simpler apparatus  
CC than column methods and result in lower loss of FP). The present sequence  
CC is that of a streptavidin binding peptide disclosed with the invention  
XX  
SQ Sequence 35 AA;

Query Match Best Local Similarity 88.2%; Score 100.5; DB 6; Length 35;  
Matches 28; Conservative 0; Mismatches 0; Indels 7; Gaps 1;

QY 1 WSHPOFEK-----XXXXXXXXXXXXXWSHPOFEK 28  
DB 1 WSHPOFEKXXXXXXXXXXXXXXXXXXXXXWSHPOFEK 35

RESULT 2  
ABP60370  
ID ABP60370 standard; peptide; 36 AA.

AC ABP60370;

DT 28-MAR-2003 (first entry)

DE Streptavidin binding peptide SEQ ID NO 11.

KW Streptavidin; protein chip; microtitre plate; detection.

OS Synthetic.

Key Location/Qualifiers

FT Region 9..28 /note= "GGGS repeats 2-5 optionally absent, residues 13-

FT 28" /label= GGGS\_repeat

PN DE10113776-A1.

PD 02-OCT-2002.

PF 21-MAR-2001; 2001DE-01013776.

PR 21-MAR-2001; 2001DE-01013776.

PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

PI Schmidt T;

PS WPI; 2003-031166/03.

PT New isolated peptide, useful as affinity purification tag for recombinant

PT protein, comprises at least two high-affinity streptavidin-binding

PT modules.

PS Claim 8; Page 16; 18pp; German.

CC The invention relates to an isolated peptide (I) comprising at least two  
CC individual modules separated by 0-50 amino acids, with each containing at  
CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain  
CC streptavidin binding modules, are useful as affinity handles for  
CC purification of recombinant fusion proteins (FP), also for detecting FP,  
CC e.g. on protein chips or microtitre plates. The modules in (I) bind  
CC strongly to streptavidin, with a co-operative effect that provides  
CC stronger binding than a single tag, but are displaced by a competitor.  
CC (I) does not interfere with the function of attached proteins (II) (so it  
CC may not be essential to remove it); facilitates detection and has easily  
CC controllable binding properties. (I) is particularly used for purifying  
CC FP from dilute solution in batch formats (which use simpler apparatus  
CC than column methods and result in lower loss of FP). The present sequence  
CC is that of a streptavidin binding peptide disclosed with the invention  
XX

SQ Sequence 36 AA;

Query Match Best Local Similarity 87.7%; Score 100; DB 6; Length 36;  
Matches 16; Conservative 12; Mismatches 0; Indels 8; Gaps 1;

QY 1 WSHPOFEK-----XXXXXXXXXXXXXWSHPOFEK 28  
DB 1 WSHPOFEKGGSGGGSGGGSGGGSGGSGWSHPOFEK 36

RESULT 3  
ABP60362  
ID ABP60362 standard; peptide; 24 AA.

AC ABP60362;

DT 28-MAR-2003 (first entry)

DE Streptavidin binding peptide SEQ ID NO 3.

KW Streptavidin; protein chip; microtitre plate; detection.

OS Synthetic.

Key Location/Qualifiers

FT Misc-difference 9..16 /label= unknown

PN DE10113776-A1.

PD 02-OCT-2002.

PF 21-MAR-2001; 2001DE-01013776.

PR 21-MAR-2001; 2001DE-01013776.

PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.

PI Schmidt T;

PS WPI; 2003-031166/03.

PT New isolated peptide, useful as affinity purification tag for recombinant

PT protein, comprises at least two high-affinity streptavidin-binding

PT modules.

PS Disclosure; Page 4; 18pp; German.

CC The invention relates to an isolated peptide (I) comprising at least two  
CC individual modules separated by 0-50 amino acids, with each containing at  
CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain  
CC streptavidin binding modules, are useful as affinity handles for  
CC purification of recombinant fusion proteins (FP), also for detecting FP,  
CC e.g. on protein chips or microtitre plates. The modules in (I) bind  
CC strongly to streptavidin, with a co-operative effect that provides  
CC stronger binding than a single tag, but are displaced by a competitor.  
CC (I) does not interfere with the function of attached proteins (II) (so it  
CC may not be essential to remove it); facilitates detection and has easily  
CC controllable binding properties. (I) is particularly used for purifying  
CC FP from dilute solution in batch formats (which use simpler apparatus  
CC than column methods and result in lower loss of FP). The present sequence  
CC is that of a streptavidin binding peptide disclosed with the invention  
XX  
SQ Sequence 24 AA;

Query Match Best Local Similarity 86.0%; Score 98; DB 6; Length 24;  
Matches 24; Conservative 0; Mismatches 0; Indels 4; Gaps 1;

QY 1 WSHPOFEKXXXXXXXXXXXXXWSHPOFEK 28  
DB 1 WSHPOFEK-----XXXXXXXXXXXXXWSHPOFEK 24

```

RESULT 4
ABP60363
ID ABP60363 standard; peptide; 24 AA.
XX
AC ABP60363;
XX
DT 28-MAR-2003 (first entry)
XX
DE Streptavidin binding peptide SEQ ID NO 4.
XX
KW Streptavidin; protein chip; microtitre plate; detection.
XX
OS Synthetic.
XX
FH Key Location/Qualifiers
FT Misc-difference 9..18
FT /label= unknown
FT /label= 22..24
FT Misc-difference 22..24
FT /label= unknown
XX
DE10113776-A1.
XX
PN 02-OCT-2002;
XX
PD 21-MAR-2001; 2001DE-01013776.
XX
PR 21-MAR-2001; 2001DE-01013776.
XX
PA (BIOA-) INST BIOANALYTIK GMBH GOETTINGEN.
XX
PI Schmidt T;
XX
DR WPI; 2003-031166/03.
XX
PT New isolated peptide, useful as affinity purification tag for recombinant
PT protein, comprises at least two high-affinity streptavidin-binding
PT modules.
XX
PS Disclosure; Page 4; 18pp; German.
XX
CC The invention relates to an isolated peptide (I) comprising at least two
CC individual modules separated by 0-50 amino acids, with each containing at
CC least one motif His-Pro-X where X = Gln, Asn or Met. (I), which contain
CC streptavidin binding modules, are useful as affinity handles for
CC purification of recombinant fusion proteins (FP), also for detecting FP,
CC e.g. on protein chips or microtitre plates. The modules in (I) bind
CC strongly to streptavidin, with a co-operative effect that provides
CC stronger binding than a single tag, but are displaced by a competitor.
CC (I) does not interfere with the function of attached proteins (II) (so it
CC may not be essential to remove it); facilitates detection and has easily
CC controllable binding properties. (I) is particularly used for purifying
CC FP from dilute solution in batch formats (which use simpler apparatus
CC than column methods and result in lower loss of FP). The present sequence
CC is that of a streptavidin binding peptide disclosed with the invention
XX
SQ Sequence 24 AA;
XX
Query Match 60.5%; Score 69; DB 6; Length 24;
Best Local Similarity 84.0%; Pred. No. 0.15;
Matches 21; Conservative 0; Mismatches 0; Indels 4; Gaps 1;
Oy 1 WSHPPFEKXXXXXXXXXXXXXWSHPQ 25
Db 1 WSHPPFEKXXXXXXXXXXXXX---HPQ 21

```

```

XX
DT 04-DEC-2003 (first entry)
XX
DE Human aggrecanase-2 (ADAMTS-5) recombinant protein sequence.
XX
KW aggrecanase; aggrecan; articular cartilage; proteoglycan;
KW cartilage compressibility; cartilage elasticity; arthritic disease;
KW osteoarthritis; cartilage degradation; inflammatory joint disease;
KW aggrecanase-1; ADAMTS-4; aggrecanase-2; ADAMTS-5; thrombospondin domain;
KW TSP domain; osteopathic; antiarthritic; cytosolic; antiinflammatory;
KW antirheumatic; ophthalmological; thrombolytic; vasotropic; antimicrobial;
KW respiratory-gen; nocotropic; neuroprotective; antiparkinsonian;
KW immunosuppressive; aggrecanase inhibition; cancer; rheumatoid arthritis;
KW septic arthritis; corneal ulceration; coronary thrombosis;
KW Crohn's disease; emphysema; Alzheimer's disease; Parkinson's disease;
KW multiple sclerosis; aortic aneurysm; enzyme; human.
XX
OS Synthetic.
XX
OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT Region 756..763
FT /label= Streptavidin_tag
XX
PN WO2003066822-A2.
XX
PD 14-AUG-2003.
XX
PF 05-FEB-2003; 2003WO-US003554.
XX
PR 05-FEB-2002; 2002US-0354592P.
XX
PA (AMHP ) WYETH.
XX
PI Georgiadis K, Crawford TK, Tomkinson KV, Morris EA, Racie LJ;
XX
DR WPI; 2003-731495/69.
XX
DR N-PSDB; ADB85491.
XX
PT New biologically-active aggrecanase protein having a deletion of all, or
PT a portion of a TSP domain, useful for treating osteoarthritis, cancer,
PT Parkinson's disease, coronary thrombosis, Alzheimer's disease and
PT multiple sclerosis.
XX
PS Disclosure; Fig 15; 11pp; English.
XX
CC This invention relates to novel truncated human aggrecanase proteins and
CC nucleotide sequences. Aggrecan is a major extracellular component of
CC articular cartilage. It is a proteoglycan responsible for providing
CC cartilage with its mechanical properties of compressibility and
CC elasticity. The loss of aggrecan has been implicated in the degradation
CC of articular cartilage in arthritic diseases such as osteoarthritis.
CC Aggrecanase is responsible for the cleavage of aggrecan, thereby having a
CC role in cartilage degradation associated with osteoarthritis and
CC inflammatory joint disease. The proteins of the current invention are
CC truncated aggrecanase-1 (ADAMTS-4) and aggrecanase-2 (ADAMTS-5) enzymes
CC which have at least one thrombospondin (TSP) domain deleted. These are
CC biologically active and have greater stability and higher expression than
CC their full-length counterparts. The proteins of the invention may be of
CC use in the development of compounds with osteopathic, antiarthritic,
CC cytostatic, antiinflammatory, antirheumatic, ophthalmological,
CC thrombolytic, vasotropic, antimicrobial, respiratory-gen, nocotropic,
CC neuroprotective, antiparkinsonian or immunosuppressive activities through
CC aggrecanase inhibition. The proteins of the invention may therefore be
CC useful for the manufacture of compositions for the treatment of
CC aggrecanase-associated conditions, such as osteoarthritis, cancer,
CC inflammatory joint disease, rheumatoid arthritis, septic arthritis,
CC corneal ulceration, coronary thrombosis, Crohn's disease, emphysema,
CC Alzheimer's disease, Parkinson's disease, multiple sclerosis and aortic
CC aneurysm. The present sequence is the amino acid sequence of a
CC recombinant truncated aggrecanase 2 (ADAMTS-5) enzyme of the invention
XX with a peptide linker and a streptavidin tag.

```

SQ Sequence 763 AA:

Query Match 59.6%; Score 68; DB 7; Length 763;  
 Best Local Similarity 42.9%; Pred. No. 37;  
 Matches 9; Conservative 12; Mismatches 0; Indels 0; Gaps 0;

QY 8 KXXXXXXXXXXMSHPQPEK 28  
 |:::|||||  
 Db 743 KGYTDVVRIPGSAMSHPOPEK 763

RESULT 6

ADS20249  
 ID ADS20249 standard; protein; 485 AA.

AC ADS20249;

DT 18-NOV-2004 (first entry)

DE Furin-processed human aggrecanase ADAMTS4 truncated protein w Strept tag.

KW ADAMTS4; a disintegrin-like and metalloprotease;  
 KW chondroitinase type I motif 4; reprotolysin; zinc metalloprotease;  
 KW aggrecanase; osteopontin; antiinflammatory; antiarthritic; antirheumatic;  
 KW cytoskeletal; osteoarthritis; glioma; cancer; inflammatory joint;  
 KW rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;  
 KW human; chromosome 1q21-q23; enzyme; truncation; mature; furin cleavage;  
 KW Strept tag.

OS Homo sapiens.  
 OS Unidentified.

FH Key Location/Qualifiers

FT Misc-difference 475..485  
 /note="Residues corresponding to positions 687-837 in the wild-type replaced by Strept tag"

PN W02004011637-A2.

PD 05-FEB-2004.

PF 29-JUL-2003; 2003WO-US023484.

PR 29-JUL-2002; 2002US-0398721P.

PA (AMHP) MYETH.

PA (CORC) CORCORAN C J.

PA (FLAN) FLANNERY C R.

PA (ZENG) ZENG W.

PA (RACI) RACIE L A.

PA (MCDO) MCDONAGH T.

PA (FREE) FREEMAN B A.

PA (GEOR) GEORGIADIS K E.

PA (LAVA) LAVALLE E R.

PI Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T;

PI Freeman BA, Georgiadis KE, Lavalie ER;

PI WPI; 2004-143860/14.

PT New isolated, modified ADAMTS4 (aggrecanase) protein with improved

PT stability useful for identifying inhibitors of the enzyme activity for

PT treating aggrecanase-associated conditions, including osteoarthritis.

PS Claim 9; SEQ ID NO 47; 117pp; English.

CC The invention relates to a novel isolated, modified ADAMTS4 (a

CC disintegrin-like and metalloprotease (reprotolysin type) with

CC chondroitinase type I motif 4) protein with improved stability compared

CC to a naturally occurring, full-length ADAMTS4 protein, where the modified

CC protein differs from the naturally-occurring, full-length ADAMTS4 protein

CC by at least one amino acid. ADAMTS proteins are a subfamily of zinc

CC metalloproteases and include aggrecanases amongst their members. The

CC protein of the invention demonstrates osteopontin, antiinflammatory,  
 CC antirheumatic, antirheumatic and cytoskeletal activities and may be useful  
 CC for treating aggrecanase-associated conditions, including osteoarthritis,  
 CC glioma, cancer, inflammatory joint disease, rheumatoid arthritis, septic  
 CC arthritis, periodontal disease and Crohn's disease. The current sequence  
 CC is that of the furin-processed human aggrecanase ADAMTS4 truncated  
 CC protein with Strept tag of the invention.

SQ Sequence 485 AA:

Query Match 59.2%; Score 67.5; DB 8; Length 485;  
 Best Local Similarity 36.7%; Pred. No. 22;  
 Matches 11; Conservative 14; Mismatches 2; Indels 3; Gaps 1;

QY 2 SHPOPEK---XXXXXXXXXXMSHPQPEK 28  
 |:::|||||  
 Db 456 SKKXFDKCMVCGDSCGSGSAMSHPOPEK 485

RESULT 7

ADS20227  
 ID ADS20227 standard; protein; 697 AA.

AC ADS20227;

DT 18-NOV-2004 (first entry)

DE Human aggrecanase ADAMTS4 truncated protein with Strept tag - SEQ ID 24.

KW ADAMTS4; a disintegrin-like and metalloprotease;  
 KW chondroitinase type I motif 4; reprotolysin; zinc metalloprotease;  
 KW aggrecanase; osteopontin; antiinflammatory; antiarthritic; antirheumatic;  
 KW cytoskeletal; osteoarthritis; glioma; cancer; inflammatory joint;  
 KW rheumatoid arthritis; septic arthritis; periodontal; Crohn's disease;  
 KW human; enzyme; chromosome 1q21-q23; truncation; Strept tag.

OS Homo sapiens.  
 OS Unidentified.

FH Key Location/Qualifiers

FT Misc-difference 687..697  
 /note="Wild-type residues 687-837 replaced by Strept tag"

PN W02004011637-A2.

PD 05-FEB-2004.

PF 29-JUL-2003; 2003WO-US023484.

PR 29-JUL-2002; 2002US-0398721P.

PA (AMHP) MYETH.

PA (CORC) CORCORAN C J.

PA (FLAN) FLANNERY C R.

PA (ZENG) ZENG W.

PA (RACI) RACIE L A.

PA (MCDO) MCDONAGH T.

PA (FREE) FREEMAN B A.

PA (GEOR) GEORGIADIS K E.

PA (LAVA) LAVALLE E R.

PI Corcoran CJ, Flannery CR, Zeng W, Racie LA, McDonagh T;

PI Freeman BA, Georgiadis KE, Lavalie ER;

PI WPI; 2004-143860/14.

PT New isolated, modified ADAMTS4 (aggrecanase) protein with improved

PT stability useful for identifying inhibitors of the enzyme activity for

PT treating aggrecanase-associated conditions, including osteoarthritis.

PS Claim 9; SEQ ID NO 24; 117pp; English.

CC The invention relates to a novel isolated, modified ADAMTS4 (a



```

PN      DE10025620-A1.
XX
PD      06-DEC-2001.
XX
PF      24-MAY-2000; 2000DE-01025620.
XX
PR      24-MAY-2000; 2000DE-01025620.
XX
PA      (GENI-) FORSCHUNGSGESELLSCHAFT GENION MBH.
XX
PI      Pongs O, Legros C, Eaucilaire M, Bougis PE,
DR      WPI; 2002-042513/06.
DR      N-PsDB; AAI69255.
XX
PT      New nucleic acid encoding chimeric potassium channel protein, useful,
PT      when immobilized on carrier, in screening agents for toxicity and
XX      pharmaceutical activity.
XX
PS      Example 1; Page 15; 34pp; German.
XX
CC      This invention describes a novel chimeric nucleic acid (I), encoding a
CC      potassium channel subunit (II) derived from Streptomyces lividans KcsA
CC      and human Kv1.3. The chimeric (bacterial-human) potassium channels are
CC      useful, after fixing to a carrier, for identifying compounds that bind to
CC      the chimerae or modify their activity, i.e. compounds that are toxic or
CC      potentially useful therapeutically, by acting as potassium channel
CC      antagonist or agonist, in treatment of circulatory or nervous system
CC      disorders, urinary incontinence, tumours, or inflammation, in human or
CC      veterinary medicine. Chimeric channels encoded by (I) can be immobilized,
CC      as an array, on a carrier, making possible screening by protein chip
CC      technology. The channels can be isolated from recombinant bacteria in
CC      stable, active and functional form, and membranes are stable for many
CC      weeks or months, during which time they can be used many times. This
CC      sequence represents Chivi subunit used in the construction of the
CC      chimeric potassium channels described in the invention
XX
SQ      Sequence 183 AA;
QY      Query Match          57.9%; Score 66; DB 5; Length 183;
        Best Local Similarity 29.2%; Pred. No. 8;
        Matches 7; Conservative 15; Mismatches 2; Indels 0; Gaps 0.
        QY      3 HQGFEXXXXXXXXXXXXXXSHPOF 26
                | :|:::|||||||
        DB      158 HERFDRLERMLDDNRSAVRHPQF 181

RESULT 10
AAG80261
ID      AAG80261 standard; protein; 183 AA.
XX
AC      AAG80261;
XX
DX      18-FEB-2002 . (first entry)
XX
DE      Chimeric potassium channel KcsA-Kv1/3 ChIVIT.
XX
KW      Potassium channel; chimeric; KcsA, Kv3.1; agonist; antagonist; tumour;
KW      circulatory disorder; nervous system disorder; urinary incontinence;
KW      inflammation; protein chip; screening; ChIVI.
XX
OS      Homo sapiens.
OS      Streptomyces lividans.
OS      Synthetic.
XX
PN      DE10025620-A1.
XX
PD      06-DEC-2001.
XX
PF      24-MAY-2000; 2000DE-01025620.
XX
PR      24-MAY-2000; 2000DE-01025620.
XX

```

```

XX XX (GENI-) FORSCHUNGSGESELLSCHAFT GENION MBH.
PA Pongs O, Legros C, Eaucilaire M, Bougis PE;
XX WPI; 2002-042513/06.
DR N-PSDB; AA169256.
XX New nucleic acid encoding chimeric potassium channel protein, useful,
PT when immobilized on carrier, in screening agents for toxicity and
XX pharmaceutical activity.
PS Example 1; Page 16; 34pp; German.
XX This invention describes a novel chimeric nucleic acid (I), encoding a
CC potassium channel subunit (II) derived from Streptomyces lividans KcsA
CC and human Kv1.3. The chimeric (bacterial-human) potassium channels are
CC useful, after fixing to a carrier, for identifying compounds that bind to
CC the chimera or modify their activity, i.e. compounds that are toxic or
CC potentially useful therapeutically, by acting as potassium channel
CC antagonist or agonist, in treatment of circulatory or nervous system
CC disorders, urinary incontinence, tumours, or inflammation, in human or
CC veterinary medicine. Chimeric channels encoded by (I) can be immobilized,
CC as an array, on a carrier, making possible screening by protein chip
CC technology. The channels can be isolated from recombinant bacteria in
CC stable, active and functional form, and membranes are stable for many
CC weeks or months, during which time they can be used many times. This
CC sequence represents the ChlVI subunit used in the construction of the
CC chimeric potassium channels described in the invention
XX Sequence 183 AA;
SQ
Query Match 57.9%; Score 66; DB 5; Length 183;
Best Local Similarity 29.2%; Pred. NO. 8;
Matches 7; Conservative 15; Mismatches 2; Indels 0; Gaps 0;
QY 3 HQPEKXXXXXXXXXXXXXWSPHPQF 26
| : ::::::::::::::: |||||
DB 158 HERFDRLERMLDDNRRSAMRHPOF 181
RESULT 11
AAG80255
ID AAG80255 standard; protein; 183 AA.
XX
AC AAG80255;
DT 18-FEB-2002 (first entry)
XX
DE Chimeric potassium channel KcsA-Kv1.3 ChII.
KW Potassium channel; chimeric; KcsA; Kv3.1; agonist; antagonist; tumour;
KW circulatory disorder; nervous system disorder; urinary incontinence;
KW inflammation; protein chip; screening; ChII.
XX Homo sapiens.
OS Streptomyces lividans.
XX Synthetic.
XX DE10025620-A1.
XX
XX 06-DEC-2001.
XX PD
PF 24-MAY-2000; 2000DE-01025620.
XX
PR 24-MAY-2000; 2000DE-01025620.
XX
(GENI-) FORSCHUNGSGESELLSCHAFT GENION MBH.
PI Pongs O, Legros C, Eaucilaire M, Bougis PE;
XX WPI; 2002-042513/06.
XX N-PSDB; AA169250.
DR

```

XX	New nucleic acid encoding chimeric potassium channel protein, useful,
PT	when immobilized on carrier, in screening agents for toxicity and
PT	pharmaceutical activity.
XX	
PS	Example 1; Page 12; 34pp; German.
XX	
CC	This invention describes a novel chimeric nucleic acid (I), encoding a
CC	potassium channel subunit (II) derived from Streptomyces lividans KcsA
CC	and human Kv1.3. The chimeric (bacterial-human) potassium channels are
CC	useful, after fixing to a carrier, for identifying compounds that bind to
CC	the chimeras or modify their activity, i.e. compounds that are toxic or
CC	potentially useful therapeutically, by acting as potassium channel
CC	antagonist or agonist, in treatment of circulatory or nervous system
CC	disorders, urinary incontinence, tumours, or inflammation, in human or
CC	veternary medicine. Chimeric channels encoded by (I) can be immobilized,
CC	as an array, on a carrier, making possible screening by protein chip
CC	technology. The channels can be isolated from recombinant bacteria in
CC	stable, active and functional form, and membranes are stable for many
CC	weeks or months, during which time they can be used many times. This
CC	sequence represents the Ch1 subunit used in the construction of the
CC	chimeric potassium channels described in the invention
XX	
SO	Sequence 183 AA;
XX	
Query Match	57.9%; Score 66; DB 5; Length 183;
Best Local Similarity	29.2%; Pred. No. 8;
Matches	7; Conservative 15; Mismatches 2; Indels 0; Gaps 0;
Oy	3 HPOFEKXXXXXXXXXXXXXWHSHPQF 26
	: : : : : : : : : : : : : : :
Db	158 HERFDRLEMLDDNRSAMRHPOF 181
XX	
RESULT 12	
AAg80258	1
ID	AAg80258 standard; protein; 183 AA.
XX	
AC	AAg80258;
XX	
DT	18-FEB-2002 (first entry)
XX	
DE	Chimeric potassium channel KcsA-Kv1/3 Ch1V.
XX	
KW	Potassium channel; chimeric; KcsA; Kv1.3; agonist; antagonist; tumour;
KW	circulatory disorder; nervous system disorder; urinary incontinence;
KW	inflammation; protein chip; screening; Ch1V.
XX	
OS	Homo sapiens.
OS	Streptomyces lividans.
OS	Synthetic.
XX	
DE	DE10025620-A1.
XX	
XX	06-DEC-2001.
XX	
PD	24-MAY-2000; 2000DE-01025620.
XX	
PR	24-MAY-2000; 2000DE-01025620.
XX	
PA	(GENI-) FORSCHUNGSGESELLSCHAFT GENION MBH.
XX	
PI	Pongs O, Legros C, Baucalire M, Bougis PE;
XX	
DR	WPI; 2002-042513/06.
DR	N-PSDB; AA169253.
XX	
PT	New nucleic acid encoding chimeric potassium channel protein, useful,
PT	when immobilized on carrier, in screening agents for toxicity and
XX	pharmaceutical activity.
XX	
PS	Example 1; Page 14; 34pp; German.
XX	

```
CC This invention describes a novel chimeric nucleic acid (I), encoding a
CC potassium channel subunit (II) derived from Streptomyces lividans KcsA
CC and human Kv1.3. The chimeric (bacterial-human) potassium channels are
CC useful, after fixing to a carrier, for identifying compounds that bind to
CC the chimeras or modify their activity, i.e. compounds that are toxic or
CC potentially useful therapeutically, by acting as potassium channel
CC antagonist or agonist, in treatment of circulatory or nervous system
CC disorders, urinary incontinence, tumours, or inflammation, in human or
CC veterinary medicine. Chimeric channels encoded by (I) can be immobilized,
CC as an array, on a carrier, making possible screening by protein chip
CC technology. The channels can be isolated from recombinant bacteria in
CC stable, active and functional form, and membranes are stable for many
CC weeks or months, during which time they can be used many times. This
CC sequence represents the ChIV subunit used in the construction of the
CC chimeric potassium channels described in the invention
SQ Sequence 183 AA;

Query Match          57.9%; Score 66; Db 5; Length 183;
Best Local Similarity 29.2%; Pred. No. 8;
Matches      7; Conservative    15; Mismatches     2; Indels      0; Gaps      0

Oy       3 HPOFEKXXXXXXXXXXWSHPQF 26
         |:|:|||||||
Db        158 HERPDLERMLDNNRSAMRHPQF 181

RESULT 13
ACAG80263
ID ACAG80263 standard; protein; 183 AA.
XX AC AG80263;
XX DT 18-FEB-2002 (first entry)
XX DE Chimeric potassium channel KcsA-Kv1.3 ChIX.
XX KM Potassium channel; chimeric; KcsA; Kv1.3; agonist; antagonist; tumour;
XX KW circulatory disorder; nervous system disorder; urinary incontinence;
XX KW inflammation; protein chip; screening; ChIX.
XX OS Homo sapiens.
XX OS Streptomyces lividans.
XX OS Synthetic.
XX FH Key location/Qualifiers
FT Misc-difference 181 /note= "Encoded by AA"
XX PN DE10025620-A1.
PD 06-DEC-2001.
XX PF 24-MAY-2000; 2000DE-01025620.
XX PR 24-MAY-2000; 2000DE-01025620.
PA (GENI-) FORSCHUNGSGESELLSCHAFT GENION MBH.
XX PI Pongs O, Legros C, Baucalre M, Bougis PE;
XX DR MPI; 2002-042513/06.
XX N-PSDB; AAI69258.
XX PT New nucleic acid encoding chimeric potassium channel protein, useful,
XX PT when immobilized on carrier, in screening agents for toxicity and
XX PT pharmaceutical activity.
XX PS Example 2; Page 17; 34pp; German.
CC This invention describes a novel chimeric nucleic acid (I), encoding a
CC potassium channel subunit (II) derived from Streptomyces lividans KcsA
CC and human Kv1.3. The chimeric (bacterial-human) potassium channels are
```







*This Page Blank (uspto)*

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2005, 12:29:14 ; Search time 87.0732 Seconds  
(without alignments)  
105.489 Million cell updates/sec

Title: SEQ16  
Perfect score: 114  
Sequence: 1 wehpqfexxxxxxxxxxxwehpqfex 28

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 1385339 seqs, 32804528 residues

Total number of hits satisfying chosen parameters: 1385339

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA:\*

1: /cgn2\_6/ptodata/2/pubpaa/US07\_PUBCOMB.pep.\*  
2: /cgn2\_6/ptodata/2/pubpaa/PCT\_NEW\_PUB.pep.\*  
3: /cgn2\_6/ptodata/2/pubpaa/US06\_NEW\_PUB.pep.\*  
4: /cgn2\_6/ptodata/2/pubpaa/US06\_PUBCOMB.pep.\*  
5: /cgn2\_6/ptodata/2/pubpaa/US07\_NEW\_PUB.pep.\*  
6: /cgn2\_6/ptodata/2/pubpaa/PCTUS\_PUBCOMB.pep.\*  
7: /cgn2\_6/ptodata/2/pubpaa/US08\_NEW\_PUB.pep.\*  
8: /cgn2\_6/ptodata/2/pubpaa/US08\_PUBCOMB.pep.\*  
9: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
10: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
11: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
12: /cgn2\_6/ptodata/2/pubpaa/US09\_PUBCOMB.pep.\*  
13: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
14: /cgn2\_6/ptodata/2/pubpaa/US10\_PUBCOMB.pep.\*  
15: /cgn2\_6/ptodata/2/pubpaa/US10C\_PUBCOMB.pep.\*  
16: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*  
17: /cgn2\_6/ptodata/2/pubpaa/US10D\_PUBCOMB.pep.\*  
18: /cgn2\_6/ptodata/2/pubpaa/US11\_NEW\_PUB.pep.\*  
19: /cgn2\_6/ptodata/2/pubpaa/US60\_NEW\_PUB.pep.\*  
20: /cgn2\_6/ptodata/2/pubpaa/US60\_PUBCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	100	87.7	36	14	US-10-026-578B-10
2	98	86.0	24	14	US-10-026-578B-3
3	87	76.3	36	14	US-10-026-578B-11
4	77	67.5	24	9	US-09-809-517A-31
5	77	67.5	25	9	US-09-809-517A-34
6	77	67.5	245	17	US-10-887-228A-1
7	77	67.5	246	17	US-10-887-228A-9
8	77	67.5	252	17	US-10-887-228A-5
9	69	60.5	24	14	US-10-026-578B-4
10	68	59.6	763	15	US-10-358-283-15
11	67.5	59.2	485	16	US-10-628-432-47
12	67.5	59.2	697	16	US-10-628-432-24
13	65	57.0	646	16	US-10-628-432-49

14	65	57.0	845	16	US-10-628-432-40	Sequence 40, Appl
15	65	57.0	858	16	US-10-628-432-27	Sequence 27, Appl
16	64.5	56.6	158	15	US-10-424-599-184769	Sequence 184769,
17	64.5	56.6	212	15	US-10-424-599-184075	Sequence 184075,
18	64	56.1	21	9	US-09-809-517A-30	Sequence 30, Appl
19	64	56.1	22	9	US-09-809-517A-33	Sequence 33, Appl
20	63.5	55.7	633	16	US-10-628-432-53	Sequence 53, Appl
21	63	55.3	117	10	US-09-977-137A-4	Sequence 4, Appl
22	63	55.3	117	10	US-09-977-137A-5	Sequence 5, Appl
23	63	55.3	117	10	US-09-977-137A-7	Sequence 7, Appl
24	63	55.3	117	10	US-09-977-137A-8	Sequence 8, Appl
25	63	55.3	117	10	US-09-977-137A-9	Sequence 9, Appl
26	63	55.3	117	10	US-09-977-137A-10	Sequence 10, Appl
27	63	55.3	117	10	US-09-977-137A-11	Sequence 11, Appl
28	63	55.3	117	10	US-09-977-137A-12	Sequence 12, Appl
29	63	55.3	118	10	US-09-977-137A-6	Sequence 6, Appl
30	63	55.3	661	15	US-10-354-983-8	Sequence 8, Appl
31	61	53.5	621	15	US-10-382-122A-50515	Sequence 50515, A
32	60.5	53.1	134	16	US-10-437-963-104065	Sequence 104065,
33	60.5	53.1	252	16	US-10-437-963-200150	Sequence 200150,
34	60	52.6	102	16	US-10-767-701-47647	Sequence 47647, A
35	60	52.6	291	15	US-10-425-114-50330	Sequence 50330, A
36	60	52.6	305	16	US-10-767-701-42846	Sequence 42846, A
37	60	52.6	309	15	US-10-425-114-61196	Sequence 61196, A
38	60	52.6	309	15	US-10-425-114-65584	Sequence 65584, A
39	60	52.6	309	15	US-10-425-114-71583	Sequence 71583, A
40	60	52.6	311	15	US-10-425-114-62543	Sequence 62543, A
41	60	52.6	311	15	US-10-425-114-70843	Sequence 70843, A
42	59	51.8	109	15	US-10-378-029-65	Sequence 65, Appl
43	59	51.8	168	16	US-10-437-963-172987	Sequence 172987,
44	59	51.8	759	16	US-10-437-963-118124	Sequence 118124,
45	59	51.8	1527	16	US-10-437-963-201843	Sequence 201843,

#### ALIGNMENTS

RESULT 1  
US-10-026-578B-10  
; Sequence 10, Application US/10026578B  
; Publication No. US20030083474A1  
; GENERAL INFORMATION:  
; APPLICANT: IBA (GmbH)  
; TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta  
; FILE REFERENCE: 100810.01US1  
; CURRENT APPLICATION NUMBER: US/10/026.578B  
; CURRENT FILING DATE: 2002-11-11  
; PRIOR APPLICATION NUMBER: DE 101 13 776.1  
; PRIOR FILING DATE: 2001-03-21  
; PRIOR APPLICATION NUMBER: PCT/EP01/11846  
; NUMBER OF SEQ ID NOS: 14  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 10  
; LENGTH: 36  
; TYPE: PRT  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Synthetic Peptide  
; FEATURE:  
; NAME/KEY: MISC FEATURE  
; LOCATION: (9)..(28)  
; OTHER INFORMATION: X represents a single amino acid at each of the positions indicat  
; OTHER INFORMATION: ed; some of the amino acids may be missing. Where amino acids are  
; FEATURE: OTHER INFORMATION: missing, the total numbers of x will be no less than 5  
; NAME/KEY: misc feature  
; OTHER INFORMATION: Artificial Sequence represents peptide binding module  
; US-10-026-578B-10  
Query Match 87.7%; Score 100; DB 14; Length 36;  
Best Local Similarity 77.8%; Pred. No. 6.6e-05;



```
Matches 10; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

Qy 5 OPEKXXXXXXXXXXWSPQPEK 28
:|||||:
Db 1 EFEOKLISEDLNGAPWSPQPEK 24

RESULT 5
US-09-809-517A-34
; Sequence 34, Application US/09809517A
; Patent No. US20020034733A1
; GENERAL INFORMATION:
; APPLICANT: Lehning, Corinna
; TITLE OF INVENTION: No. US20020034733A1, methods for displaying (poly)peptides/prote
; FILE REFERENCE: MORPHO/11
; CURRENT FILING DATE: 2001-03-15
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 34
; LENGTH: 25
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34

Query Match 67.5%; Score 77; DB 9; Length 25;
Best Local Similarity 41.7%; Pred. No. 0.035;
Matches 10; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

Qy 5 OPEKXXXXXXXXXXWSPQPEK 28
:|||||:
Db 2 EFEOKLISEDLNGAPWSPQPEK 25

RESULT 6
US-10-887-228A-1
; Sequence 1, Application US/10887228A
; Publication No. US20050037402A1
; GENERAL INFORMATION:
; APPLICANT: Schering AG
; TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio
; FILE REFERENCE: S30569US
; CURRENT FILING DATE: 2004-07-09
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: DE 103 31 093.2
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US 60/478,262
; PRIOR FILING DATE: 2003-07-16
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 1
; LENGTH: 245
; TYPE: PRT
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-1

Query Match 67.5%; Score 77; DB 17; Length 245;
Best Local Similarity 41.7%; Pred. No. 0.74;
Matches 10; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

Qy 5 OPEKXXXXXXXXXXWSPQPEK 28
:|||||:
Db 222 EFEOKLISEDLNGAPWSPQPEK 245
```

```
RESULT 7
US-10-887-228A-9
; Sequence 9, Application US/10887228A
; Publication No. US20050037402A1
; GENERAL INFORMATION:
; APPLICANT: Schering AG
; TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio
; FILE REFERENCE: S30569US
; CURRENT FILING DATE: 2004-07-09
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: DE 103 31 093.2
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US 60/478,262
; PRIOR FILING DATE: 2003-07-16
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 9
; LENGTH: 246
; TYPE: PRT
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-9

Query Match 67.5%; Score 77; DB 17; Length 246;
Best Local Similarity 41.7%; Pred. No. 0.74;
Matches 10; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

Qy 5 OPEKXXXXXXXXXXWSPQPEK 28
:|||||:
Db 223 EFEOKLISEDLNGAPWSPQPEK 246

RESULT 8
US-10-887-228A-5
; Sequence 5, Application US/10887228A
; Publication No. US20050037402A1
; GENERAL INFORMATION:
; APPLICANT: Schering AG
; TITLE OF INVENTION: Device and Process for Direct Quantitative in vitro Determinatio
; FILE REFERENCE: S30569US
; CURRENT FILING DATE: 2004-07-09
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: DE 103 31 093.2
; PRIOR FILING DATE: 2003-07-09
; PRIOR APPLICATION NUMBER: US 60/478,262
; PRIOR FILING DATE: 2003-07-16
; NUMBER OF SEQ ID NOS: 12
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 5
; LENGTH: 252
; TYPE: PRT
; ORGANISM: artificial
; FEATURE:
; OTHER INFORMATION: Fab fragment of variable heavy chain
US-10-887-228A-5

Query Match 67.5%; Score 77; DB 17; Length 252;
Best Local Similarity 41.7%; Pred. No. 0.77;
Matches 10; Conservative 14; Mismatches 0; Indels 0; Gaps 0;

Qy 5 OPEKXXXXXXXXXXWSPQPEK 28
:|||||:
Db 229 EFEOKLISEDLNGAPWSPQPEK 252

RESULT 9
US-10-026-578B-4
; Sequence 4, Application US/10026578B
```

```

Publication No. US20030083474A1
GENERAL INFORMATION:
APPLICANT: IBA (GmbH)
APPLICANT: Schmidt, Thomas
TITLE OF INVENTION: Sequentially Arranged Streptavidin-Binding Modules as Affinity Ta
FILE REFERENCE: 100810.01US1
CURRENT APPLICATION NUMBER: US/10/026,578B
PRIOR FILING DATE: 2002-11-11
PRIOR APPLICATION NUMBER: DE 101 13 776.1
PRIOR FILING DATE: 2001-03-21
PRIOR APPLICATION NUMBER: PCT/EP01/11846
PRIOR FILING DATE: 2001-10-12
NUMBER OF SEQ ID NOS: 14
SOFTWARE: PatentIn version 3.1
SEQ ID NO 4
LENGTH: 24
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Synthetic Peptide
NAME/KEY: MISC_FEATURE
LOCATION: (9)..(9)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (10)..(10)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (11)..(11)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (12)..(12)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (13)..(13)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (14)..(14)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (15)..(15)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (16)..(16)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (17)..(17)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (18)..(18)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (22)..(22)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (23)..(23)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:
NAME/KEY: MISC_FEATURE
LOCATION: (24)..(24)
OTHER INFORMATION: X represents a single amino acid at the position indicated
FEATURE:

```

```

NAME/KEY: misc feature
OTHER INFORMATION: Artificial Sequence represents peptide binding module
US-10-026-578B-4

```

```

Query Match 60.5%; Score 69; DB 14; Length 24;
Best Local Similarity 84.0%; Pred. No. 0.34;
Matches 21; Conservative 0; Mismatches 0; Indels 4; Gaps 1;
QY 1 WSHPOFEKXXXXXXXXXXWSHPQ 25
DB 1 WSHPOFEKXXXXXXXXXX----HPQ 21

```

```

RESULT 10
US-10-358-283-15
Sequence 15, Application US/10358283
Publication No. US20040054149A1
GENERAL INFORMATION:
APPLICANT: WYETH
TITLE OF INVENTION: TRUNCATED AGGREGINASE MOLECULES
FILE REFERENCE: 08702-0112-00000
CURRENT APPLICATION NUMBER: US/10/358,283
PRIOR FILING DATE: 2003-02-17
PRIOR APPLICATION NUMBER: 60/354,592
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PatentIn Ver. 2.1
SEQ ID NO 15
LENGTH: 763
TYPE: PRT
ORGANISM: Homo sapiens
US-10-358-283-15

```

```

Query Match 59.6%; Score 68; DB 15; Length 763;
Best Local Similarity 42.9%; Pred. No. 48;
Matches 9; Conservative 12; Mismatches 0; Indels 0; Gaps 0;
QY 8 KXXXXXXXXXXWSHPQFEK 28
DB 743 KGYTDVVRIPGSAWSHPQFEK 763

```

```

RESULT 11
US-10-628-432-47
Sequence 47, Application US/10628432
Publication No. US20040142863A1
GENERAL INFORMATION:
APPLICANT: Wyeth
TITLE OF INVENTION: Modified ADAMTS4 molecules
FILE REFERENCE: AM101378
CURRENT APPLICATION NUMBER: US/10/628,432
PRIOR FILING DATE: 2003-07-29
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.1
SEQ ID NO 47
LENGTH: 485
TYPE: PRT
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: furin-processed construct C
US-10-628-432-47

```

```

Query Match 59.2%; Score 67.5; DB 16; Length 485;
Best Local Similarity 36.7%; Pred. No. 30;
Matches 11; Conservative 14; Mismatches 2; Indels 3; Gaps 1;
QY 2 SHPOFEK---XXXXXXXXXXWSHPQFEK 28
DB 456 SKKFDKCMVCGSDSGSGSAWSHPQFEK 485

```

```

RESULT 12
US-10-628-432-24

```

```

Sequence 24, Application US/10628432
Publication No. US20040142863A1
GENERAL INFORMATION:
APPLICANT: Wyeth
TITLE OF INVENTION: Modified ADAMTS4 molecules
FILE REFERENCE: AM101378
CURRENT APPLICATION NUMBER: US/10/628,432
CURRENT FILING DATE: 2003-07-29
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.1
SEQ ID NO 24
LENGTH: 697
TYPE: PRT
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: Truncated ADAMTS4 molecule
US-10-628-432-24

Query Match          59.2% Score 67.5; DB 16; Length 697;
Best Local Similarity 36.7%; Pred. No. 49;
Matches 11; Conservative 14; Mismatches 2; Indels 3; Gaps 1

Oy      2 SHPOFEK---XXXXXXXXXXXXXSHPOFEK 28
       ||||| : : : : : : : : : : : : : :
Db      668 SKKRPDKMVGCGDGGSGGSMSHPOFEK 697


RESULT 13
US-10-628-432-49,
Sequence 49, Application US/10628432
Publication No. US20040142863A1
GENERAL INFORMATION:
APPLICANT: Wyeth
TITLE OF INVENTION: Modified ADAMTS4 molecules
FILE REFERENCE: AM101378
CURRENT APPLICATION NUMBER: US/10/628,432
CURRENT FILING DATE: 2003-07-29
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.1
SEQ ID NO 49
LENGTH: 646
TYPE: PRT
ORGANISM: Artificial
FEATURE:
OTHER INFORMATION: furlin-processed construct E
US-10-628-432-49

Query Match          57.0% Score 65; DB 16; Length 646;
Best Local Similarity 31.2%; Pred. No. 92;
Matches 10; Conservative 14; Mismatches 4; Indels 4; Gaps 1

Oy      1 WSHPOFE-----KXXXXXXXXXXXXXSHPOFEK 28
       ||||| : : : : : : : : : : : : : :
Db      615 WLHRRQAQLLEILRRPFWAGRGKSAMSHPOFEK 646


RESULT 14
US-10-628-432-40
Sequence 40, Application US/10628432
Publication No. US20040142863A1
GENERAL INFORMATION:
APPLICANT: Wyeth
TITLE OF INVENTION: Modified ADAMTS4 molecules
FILE REFERENCE: AM101378
CURRENT APPLICATION NUMBER: US/10/628,432
CURRENT FILING DATE: 2003-07-29
NUMBER OF SEQ ID NOS: 53
SOFTWARE: PatentIn version 3.1
SEQ ID NO 40
LENGTH: 845
TYPE: PRT
ORGANISM: Artificial
FEATURE:
```

```

: OTHER INFORMATION: ADAMTS4 ASM with insertion
US-10-628-432-40

Query Match          57.0%: Score 65; DB 16; Length 845;
Best Local Similarity 40.9%: Pred. No. 1.3e+02;
Matches 9; Conservative 12; Mismatches 1; Indels 0; Gaps 0;

Oy      1 WSHPOFEKXXXXXXXXXWS 22
      |||||:|||||:
Db      520 WSHPOFEKAGCGPWPWGDCS 541

RESULT 15
US-10-628-432-27
: Sequence 27, Application US/10628432
: Publication No. US20040142863A1
: GENERAL INFORMATION:
: APPLICANT: Wyeth
: TITLE OF INVENTION: Modified ADAMTS4 molecules
: FILE REFERENCE: AM101378
: CURRENT APPLICATION NUMBER: US/10/628,432
: CURRENT FILING DATE: 2003-07-29
: NUMBER OF SEQ ID NOS: 53
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 27
: LENGTH: 858
: TYPE: PRT
: ORGANISM: Artificial
: FEATURE:
: OTHER INFORMATION: modified ADAMTS4 molecule
US-10-628-432-27

Query Match          57.0%: Score 65; DB 16; Length 858;
Best Local Similarity 31.2%: Pred. No. 1.3e+02;
Matches 10; Conservative 14; Mismatches 4; Indels 4; Gaps 1;

Oy      1 WSHPOFE---KXXXXXXXXXXWSHPQFEK 28
      |||:|:|||||:|||||
Db      827 WLMHRAQLLEILRRRPPWAGRGSAWSHPQFEK 858

Search completed: March 2, 2005, 14:19:04
Job time : 88.0732 secs

```

*This Page Blank (uspto)*



GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

## OM protein - protein search, using sw model

Run on: March 2, 2005, 12:20:44 ; Search time 31.8699 Seconds  
(without alignments)  
65.585 Million cell updates/sec

Title: SEQ16

Perfect score: 114  
Sequence: 1 webpqrfehxkxxxxxxxxxxwebpqrfehxk 28

Scoring table: BLOSUM62DX  
Gapop 10.0 , Gapext 0.5

Searched: 513545 seqs, 74649064 residues

Total number of hits satisfying chosen parameters: 513545

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents, AA:\*

- 1: /cgn2\_6/ptodata/1/1aa/5A\_COMB.pep:\*
- 2: /cgn2\_6/ptodata/1/1aa/5B\_COMB.pep:\*
- 3: /cgn2\_6/ptodata/1/1aa/6A\_COMB.pep:\*
- 4: /cgn2\_6/ptodata/1/1aa/6B\_COMB.pep:\*
- 5: /cgn2\_6/ptodata/1/1aa/PTOS\_COMB.pep:\*
- 6: /cgn2\_6/ptodata/1/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	77	67.5	24	US-09-809-517A-31	Sequence 31, Appl
2	77	67.5	25	US-09-809-517A-34	Sequence 34, Appl
3	64	56.1	21	US-09-809-517A-30	Sequence 30, Appl
4	64	56.1	22	US-09-809-517A-33	Sequence 33, Appl
5	63	55.3	117	US-09-977-137A-4	Sequence 4, Appl1
6	63	55.3	117	US-09-977-137A-5	Sequence 5, Appl1
7	63	55.3	117	US-09-977-137A-7	Sequence 7, Appl1
8	63	55.3	117	US-09-977-137A-8	Sequence 8, Appl1
9	63	55.3	117	US-09-977-137A-9	Sequence 9, Appl1
10	63	55.3	117	US-09-977-137A-10	Sequence 10, Appl1
11	63	55.3	117	US-09-977-137A-11	Sequence 11, Appl1
12	63	55.3	117	US-09-977-137A-12	Sequence 12, Appl1
13	63	55.3	118	US-09-977-137A-6	Sequence 6, Appl1
14	58	50.9	129	US-09-248-796A-22626	Sequence 22626, A
15	57	50.9	129	US-09-270-767-31955	Sequence 31955, A
16	57	50.0	168	US-09-270-767-47172	Sequence 47172, A
17	56	48.1	389	US-08-811-949-65	Sequence 481, Appl
18	56	48.1	389	US-09-538-092-481	Sequence 481, App
19	56	49.1	829	US-09-538-092-995	Sequence 995, App
20	56	49.1	829	US-09-252-991A-24768	Sequence 24768, A
21	55	48.2	138	US-09-902-540-15911	Sequence 15911, A
22	55	48.2	252	US-09-976-594-454	Sequence 454, App
23	55	48.2	260	US-09-270-767-32978	Sequence 32978, A
24	55	48.2	260	US-09-270-767-48195	Sequence 48195, A
25	55	48.2	288	US-09-248-796A-24741	Sequence 24741, A
26	54.5	47.8	1708	US-09-462-606-2	Sequence 2, Appl1
27	54	47.4	109	US-09-902-540-14610	Sequence 14610, A

28	54	47.4	259	4	US-09-431-887-33	Sequence 33, Appl
29	53.5	46.9	155	3	US-08-828-741B-11	Sequence 11, Appl
30	53.5	46.9	155	3	US-09-160-567-11	Sequence 11, Appl
31	53.5	46.9	155	4	US-09-710-299-11	Sequence 11, Appl
32	53.5	46.9	155	4	US-09-509-031-11	Sequence 11, Appl
33	53.5	46.9	495	3	US-08-828-741B-4	Sequence 4, Appl1
34	53.5	46.9	495	3	US-09-160-567-4	Sequence 4, Appl1
35	53.5	46.9	495	4	US-09-710-299-4	Sequence 4, Appl1
36	53.5	46.9	495	4	US-09-509-031-4	Sequence 4, Appl1
37	53	46.5	10	4	US-09-809-517A-6	Sequence 6, Appl1
38	53	46.5	203	4	US-09-270-767-33748	Sequence 33748, A
39	53	46.5	297	4	US-09-538-092-546	Sequence 546, App
40	53	46.5	410	4	US-09-107-532A-6066	Sequence 6066, App
41	53	46.5	482	4	US-09-107-532A-4868	Sequence 4868, App
42	53	46.5	482	4	US-09-248-796A-18158	Sequence 18158, A
43	53	46.5	527	4	US-09-600-985-1	Sequence 1, Appl1
44	53	46.5	527	4	US-09-600-985-2	Sequence 2, Appl1
45	53	46.5	527	4	US-09-600-985-3	Sequence 3, Appl1

## ALIGNMENTS

```
RESULT 1
US-09-809-517A-31
; Sequence 31, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on t
; FILE OF INVENTION: Particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
; SOFTWARE: Patent version 3.0
; SEQ ID NO 31
; LENGTH: 24
; TYPE: PRT
; ORGANISM: artificial sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-31
Query Match          67.5%; Score 77; DB 4; Length 24;
Best Local Similarity 41.7%; Pred.No. 0.0013;
Matches 10; Conservative 14; Mismatches 0; Indels 0; Gaps 0;
QY 5 OPEKXXXXXXXXXXWSHPQEK 28
:|||||:
DB 1 EPEKILSEBDNGAPWSHPQEK 24
RESULT 2
US-09-809-517A-34
; Sequence 34, Application US/09809517A
; Patent No. 6753136
; GENERAL INFORMATION:
; APPLICANT: Lohning, Corinna
; TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on t
; FILE OF INVENTION: Particles via disulfide bonds
; FILE REFERENCE: MORPHO/11
; CURRENT APPLICATION NUMBER: US/09/809,517A
; PRIOR FILING DATE: 2001-03-15
; PRIOR APPLICATION NUMBER: EP 99114072.4
; PRIOR FILING DATE: 1999-07-20
; PRIOR APPLICATION NUMBER: EP 00103551.8
; PRIOR FILING DATE: 2000-02-18
; NUMBER OF SEQ ID NOS: 41
```

```
/ SOFTWARE: Patentin version 3.0
/ SEQ ID NO 34
/ LENGTH: 25
/ TYPE: PRT
/ ORGANISM: artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-34
```

```
Query Match          67.5%; Score 77; DB 4; Length 25;
Best Local Similarity 41.7%; Pred. No. 0.0014;
Matches 10; Conservative 14; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 5 OXXXXXXXXXXXXXXXXXSHPOPEK 28
Db 2 EFDYKDDDDKGAQWMSHPQPEK 25
```

```
RESULT 3
US-09-809-517A-30
/ Sequence 30, Application US/09809517A
/ Patent No. 6753136
/ GENERAL INFORMATION:
/ APPLICANT: Lohning, Corinna
/ TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on h
/ FILE REFERENCE: MORPHO/11
/ CURRENT APPLICATION NUMBER: US/09/809,517A
/ PRIOR FILING DATE: 2001-03-15
/ PRIOR APPLICATION NUMBER: EP 99114072.4
/ PRIOR FILING DATE: 1999-07-20
/ PRIOR APPLICATION NUMBER: EP 00103551.8
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: Patentin version 3.0
/ SEQ ID NO 30
/ LENGTH: 21
/ TYPE: PRT
/ ORGANISM: artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-30
```

```
Query Match          56.1%; Score 64; DB 4; Length 21;
Best Local Similarity 38.1%; Pred. No. 0.075;
Matches 8; Conservative 13; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 8 KXXXXXXXXXXXXXXXXXSHPOPEK 28
Db 1 EFDYKDDDDKGAQWMSHPQPEK 21
```

```
RESULT 4
US-09-809-517A-33
/ Sequence 33, Application US/09809517A
/ Patent No. 6753136
/ GENERAL INFORMATION:
/ APPLICANT: Lohning, Corinna
/ TITLE OF INVENTION: No. 6753136el methods for displaying (poly)peptides/proteins on h
/ FILE REFERENCE: MORPHO/11
/ CURRENT APPLICATION NUMBER: US/09/809,517A
/ PRIOR FILING DATE: 2001-03-15
/ PRIOR APPLICATION NUMBER: EP 99114072.4
/ PRIOR FILING DATE: 1999-07-20
/ PRIOR APPLICATION NUMBER: EP 00103551.8
/ NUMBER OF SEQ ID NOS: 41
/ SOFTWARE: Patentin version 3.0
/ SEQ ID NO 33
/ LENGTH: 22
/ TYPE: PRT
/ ORGANISM: artificial sequence
```

```
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: synthetic module
US-09-809-517A-33
```

```
Query Match          56.1%; Score 64; DB 4; Length 22;
Best Local Similarity 38.1%; Pred. No. 0.081;
Matches 8; Conservative 13; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 8 KXXXXXXXXXXXXXXXXXSHPOPEK 28
Db 2 EFDYKDDDDKGAQWMSHPQPEK 22
```

```
RESULT 5
US-09-977-137A-4
/ Sequence 4, Application US/09977137A
/ Patent No. 6750042
/ GENERAL INFORMATION:
/ APPLICANT: Summers, Anne O.
/ TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
/ FILE REFERENCE: 79-00
/ CURRENT APPLICATION NUMBER: US/09/977,137A
/ PRIOR FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,465
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 4
/ LENGTH: 117
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-4
```

```
Query Match          55.3%; Score 63; DB 4; Length 117;
Best Local Similarity 40.0%; Pred. No. 1.5;
Matches 8; Conservative 12; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 9 XXXXXXXXXXXXXXXXXXXXSHPOPEK 28
Db 98 HARKGNVSCPSAMSHQPEK 117
```

```
RESULT 6
US-09-977-137A-5
/ Sequence 5, Application US/09977137A
/ Patent No. 6750042
/ GENERAL INFORMATION:
/ APPLICANT: Summers, Anne O.
/ TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
/ FILE REFERENCE: 79-00
/ CURRENT APPLICATION NUMBER: US/09/977,137A
/ PRIOR FILING DATE: 2001-10-12
/ PRIOR APPLICATION NUMBER: US 60/240,465
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: Patentin Ver. 2.0
/ SEQ ID NO 5
/ LENGTH: 117
/ TYPE: PRT
/ ORGANISM: Artificial Sequence
/ FEATURE:
/ OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-5
```

```
Query Match          55.3%; Score 63; DB 4; Length 117;
Best Local Similarity 40.0%; Pred. No. 1.5;
Matches 8; Conservative 12; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      9 XXXXXXXXXXXXWSPQFEK 28
          :::::::::::|||||
Db      98 HARKGNVSCPSAWSPQFEK 117
```

```

RESULT 7
US-09-977-137A-7
Sequence 7, Application US/09977137A
Patent No. 6750042
GENERAL INFORMATION:
APPLICANT: Summers, Anne O.
APPLICANT: Caguiat, Jonathan
TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
TITLE OF INVENTION: Methods
FILE REFERENCE: 79-00
CURRENT APPLICATION NUMBER: US/09/977,137A
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,465
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 7
LENGTH: 117
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-7

```

```

RESULT 8
US-09-977-137A-8
Sequence 8, Application US/09977137A
Patent No. 6750042
GENERAL INFORMATION:
APPLICANT: Summers, Anne O.
APPLICANT: Cegulat, Jonathan
TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
FILE REFERENCE: 79-00
CURRENT APPLICATION NUMBER: US/09/977,137A
CURRENT FILING DATE: 2001-10-12
PRIOR APPLICATION NUMBER: US 60/240,465
PRIOR FILING DATE: 2000-10-12
NUMBER OF SEQ ID NOS: 18
SOFTWARE: Patentin Ver. 2.0
SEQ ID NO 8
LENGTH: 117
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURES:
OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-8

```

	Query Match	Score 63;	DB 4;	length 117;
	Best Local Similarity	40.0%;	Pred. No. 1.5;	
Matches	8; Conservative	12; Mismatches	0; Indels	0; Gaps
Oy	9 xxxxxxxxxxxxwshpqfex	28 :::~::~:		
Dd	98 HARKGNVSCPSAWSHQFEX	117       		

```

US-09-977-137A-9
; Sequence 9, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summers, Anne O.
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 9
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
US-09-977-137A-9

Query Match          55.3%; Score 63; DB 4; Length 117;
Best Local Similarity 40.0%; Pred.No. 1.5;
Matches      8; Conservative 12; Mismatches      0; Indels      0; Gaps      0;

QY      9 XXXXXXXXXXXXXGSHPOPEK 28
          ::::::::::::::|
          98 HARKGNVSCPSAWSHPOPEK 117

```

```

RESULT 10
; US-09-977-137A-10
; Sequence 10, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summeys, Anne O.
; APPLICANT: Caguiat, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and
; TITLE OF INVENTION: Methods
; FILE REFERENCE: 79-00
; CURRENT APPLICATION NUMBER: US/09/977,137A
; CURRENT FILING DATE: 2001-10-12
; PRIOR APPLICATION NUMBER: US 60/240,465
; PRIOR FILING DATE: 2000-10-12
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 117
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: chelon
; US-09-977-137A-10

```

```

RESULT 11
US-09-977-137A-11
; Sequence 11, Application US/09977137A
; Patent No. 6750042
; GENERAL INFORMATION:
; APPLICANT: Summeys, Anne O.
; INVENTOR: Cagliati, Jonathan
; TITLE OF INVENTION: Metal Binding Proteins, Recombinant Host Cells and

```

```

Query Match      50.0%; Score 57; DB 4; Length 168;
Best Local Similarity 26.9%; Pred. No. 19;
Matches 7; Conservative 14; Mismatches 5; Indels 0; Gaps 0.

QY      1 WSHPOFEKXXXXXXXXXXXXXWSHPQF 26
      : ||| | : : : : : : : : : : : :

```

Db 14 YKHPQETKSNRRDRAGRLISEWARP1F 39

Search completed: March 2, 2005, 12:25:39  
Job time : 31.8699 secs

*This Page Blank (uspto)*